Mind in Life

BIOLOGY, PHENOMENOLOGY,
AND THE SCIENCES OF MIND

◆ ◆ ◆

Evan Thompson

THE BELKNAP PRESS OF
HARVARD UNIVERSITY PRESS
Cambridge, Massachusetts
London, England

Copyright © 2007 by the President and Fellows of Harvard College
All rights reserved
Printed in the United States of America

First Harvard University Press paperback edition, 2010

Library of Congress Cataloging-in-Publication Data

Thompson, Evan.
Mind in life : biology, phenomenology, and the sciences of mind / Evan Thompson.
p. cm.
Includes bibliographical references and index.
ISBN 978–0–674–02511–0 (cloth : alk. paper)
ISBN 978-0–674–05751–7 (pbk.)
1. Consciousness. 2. Philosophy of mind. 3. Phenomenology. 4. Mind and body.
5. Experience. I. Title.
B808.9.T45 2007
128'.2—dc22 2006050713

To

Gabriel Cohen Varela
Maximilian Todd Williams
Gareth Todd Thompson

Contents

PART THREE Consciousness in Life *219*

Preface

THE THEME OF THIS BOOK is the deep continuity of life and mind. Where there is life there is mind, and mind in its most articulated forms belongs to life. Life and mind share a core set of formal or organizational properties, and the formal or organizational properties distinctive of mind are an enriched version of those fundamental to life. More precisely, the *self-organizing* features of mind are an enriched version of the self-organizing features of life. The self-producing or "autopoietic" organization of biological life already implies cognition, and this incipient mind finds sentient expression in the self-organizing dynamics of action, perception, and emotion, as well as in the self-moving flow of time-consciousness.

From this perspective, mental life is also bodily life and is situated in the world. The roots of mental life lie not simply in the brain, but ramify through the body and environment. Our mental lives involve our body and the world beyond the surface membrane of our organism, and therefore cannot be reduced simply to brain processes inside the head.

The chapters to come elaborate these ideas using material drawn from three main sources—biology, phenomenological philosophy, and psychology and neuroscience. The book as a whole is intended to bring the experimental sciences of life and mind into a closer and more harmonious relationship with phenomenological investigations of experience and subjectivity.

The principal motive behind this aim is to make headway on one of the outstanding philosophical and scientific problems of our time—the so-called explanatory gap between consciousness and nature. Exactly how are consciousness and subjective experience related to the brain and body? It is one thing to be able to establish correlations between consciousness and brain activity; it is another thing to have an account that explains exactly how certain biological processes generate and realize consciousness and subjectivity. At the present time, we not only lack such an account, but also are unsure about the form it would need to have in order to bridge the conceptual and epistemological gap between life and mind as objects of scientific investigation, and life and mind as we subjectively experience them.

In this book, I offer no new or original theory or model of consciousness, no new conceptual analysis of physical and phenomenal concepts, and no new speculative metaphysical synthesis to unify consciousness and nature. My aim and approach are different. To make real progress on the explanatory gap, we need richer phenomenological accounts of the structure of experience, and we need scientific accounts of mind and life informed by these phenomenological accounts. Phenomenology in turn needs to be informed by psychology, neuroscience, and biology. My aim is not to close the explanatory gap in a reductive sense, but rather to enlarge and enrich the philosophical and scientific resources we have for addressing the gap. My approach is thus to bring phenomenological analyses of experience into a mutually illuminating relationship with scientific analyses of life and mind.

Acknowledgments

I CANNOT HELP but think of this book as exemplifying what philosophers call the bundle theory of personal identity. According to the bundle theory, there is no single and permanent self that persists through time; the self is rather a bundle of constantly changing and psychologically continuous experiences or mental episodes. Similarly, this book began many years ago and has undergone so many transformations since its inception that I cannot say with any confidence that it is the "same" book I started work on more than ten years ago.

Originally, this book (or its textual ancestor) was supposed to be co-authored with Francisco Varela. We had hoped to write a follow-up to our book (co-authored with Eleanor Rosch), *The Embodied Mind: Cognitive Science and Human Experience* (MIT Press, 1991). When we began planning our new book (in 1994), Francisco had just learned that he was chronically ill with Hepatitis C. Thus, from the beginning, a sense of urgency lay over this book. Eventually it became clear that Francisco would need a liver transplant. At this time (in 1998), Francisco decided to step back from the project and encouraged me to continue on my own. I thus set about to revise the book by myself. After the success of the transplant, Francisco felt new enthusiasm for the project, and we tried to resume our collaborative efforts. Sadly, his illness returned not long afterward, and Francisco died on May 28, 2001, at his home in Paris. The obituary I wrote a few days later for the online journal *Psyche* can be read at: http://psyche.cs.monash.edu.au/v7/

psyche-7–12-thompson.html An abridged version was also published in *Journal of Consciousness Studies* 8 (2001): 66–69.

After Francisco's death, I tried to continue writing this book as a coauthored one. But there was still a large amount of writing to be done, and as time passed it became clear that the book needed to be completely recast and rewritten by me alone. I reorganized and rewrote the chapters, and changed the title twice, before the book finally took its present form. Thus, although the enormous influence of Francisco's thought will be evident to anyone who reads this book, I bear full responsibility for this work's contents, and all shortcomings and errors are mine.

Over the long and difficult time it has taken to produce this book, I have had the support and encouragement of many people.

No one has given me more encouragement, aid, and love, and shared more in the tumultuous life of this book, than Rebecca Todd. The love and gratitude I feel are beyond expression. In addition to living with my long periods of self-absorption while struggling with this book, she has read numerous drafts and helped me improve my thinking and writing immeasurably. I cannot imagine having written this book without her.

Our sons Maximilian Todd Williams and Gareth Todd Thompson have had to endure my working on this book for almost their entire lives. My gratitude to them is boundless. I dedicate this book to them, and to our dear family friend, Gabriel Cohen Varela, son of Francisco Varela and Amy Cohen Varela, and the same age as my younger son.

Gail Thompson, William Irwin Thompson, and Hilary Thompson have given great support in ways too numerous to detail. I am deeply grateful to them.

I also express deep thanks to John and Nancy Todd for their encouragement over many years.

Amy Cohen Varela's enthusiasm for this book, and her unhesitating support and encouragement when I decided to transform it into a single-authored work after Francisco's death, are deeply appreciated.

Alva Noë's friendship and intellectual companionship during the writing of this book are greatly valued. I express my gratitude to him also for comments on early versions of several chapters, and for his reading and commenting on the penultimate version of the entire manuscript.

Many people over the years have read various portions of this book and given helpful advice along the way: Giovanna Colombetti, Diego Cosmelli, Isabela Granic, Robert Hanna, Marc Lewis, Antoine Lutz, Susan Oyama, Michael Silberstein, Dan Zahavi, and Phil Zelazo.

For helpful discussions on a wide variety of topics related to this book, I thank Lisa Adams, Ralph Adolphs, Michel Bitbol, Maria Botero, Paul Bourgine, Richie Davidson, Natalie Depraz, Jean-Pierre Dupuy, Frans de Waal, Adam Engle, Michel Ferrari, Shaun Gallagher, Vittorio Gallese, Tom Hollenstein, Evelyn Fox Keller, Uriah Kriegel, Jean-Philippe Lachaux, Connie Lamm, Alex Lamey, Dorothée Legrand, Michel Le Van Quyen, Luigi Luisi, Jun Luo, Eduard Marbach, Barry McMullin, Bernard Pachoud, Slobodan Perovic, Claire Petitmengin, Jean Petitot, Ljiljana Radenovic, Franz Reichle, Eugenio Rodriguez, Andreas Roepstorff, Jean-Michel Roy, David Rudrauf, Brian Cantwell Smith, John Stewart, Steve Torrence, Joel Walmsley, Mark Wexler, and Arthur Zajonc.

The penultimate version of this book served as the main text for a graduate seminar in "Phenomenological Philosophy of Mind" I taught during the Fall Term 2005, in the Department of Philosophy at the University of Toronto. I am grateful to everyone who attended that seminar—especially Ranpal Dosanjh, David Egan, Cathal Ó Madagáin, Joshua Ben Nichols, Adrienne Prettyman, and Joel Walmsley—for their critical responses to the text.

Helena De Preester, at short notice, read the penultimate version of the text and sent me numerous detailed and insightful comments that helped my final revisions considerably.

Mike Wheeler also read the entire manuscript and helpfully called attention to several places where my arguments could be improved.

I am particularly thankful for the patience of Lindsay Waters at the Harvard University Press. His encouragement and speedy efficiency when the manuscript finally arrived on his desk are much appreciated.

I am grateful to several institutions and individuals for their support. During March and April 2003, I stayed as a visting Mâitre de Recherche at the Centre de Recherche en Epistémologie Appliqué (CREA), at the Ecole Polytechnique, Paris. I wish to thank its director, Jean Petitot, for this opportunity, which greatly aided me in writing this book. I also thank Dan Zahavi, director of the Center for Subjectivity Research at the University of Copenhagen, for inviting me to visit for two weeks in

June 2004 and June 2005, and for numerous productive exchanges during this time. Finally, I am grateful to the McDonnell Project in Philosophy and the Neurosciences, its director, Kathleen Akins, and my fellow project members, for their interest in and support of my work.

For financial support, I am grateful to the Social Sciences and Humanities Research Council of Canada for a Standard Research Grant (1998–2001) and a Canada Research Chair (2002–2007); York University for a Faculty of Arts Research Fellowship (2001–2002); the McDonnell Project in Philosophy and Neuroscience; the Center for Consciousness Studies at the University of Arizona, Tucson; and the University of Toronto for a Connaught Start-up Research Award.

An earlier version of Chapter 9 was published as the article "Sensorimotor Subjectivity and the Enactive Approach to Experience," in *Phenomenology and the Cognitive Sciences* 4 (2005): 407–427 (© Springer-Verlag). Chapter 10 is an expanded version of "Look Again: Phenomenology and Mental Imagery," in *Phenomenology and the Cognitive Sciences* (in press; © Springer-Verlag), and of "Representationalism and the Phenomenology of Mental Imagery," in *Synthese* (in press; © Springer-Verlag). I am grateful to the editor of these journals for permission to use this material in revised form here.

Finally, I am grateful to my former colleagues at York University and my new colleagues at the University of Toronto for their encouragement and interest in my work.

PART ONE

◆

The Enactive Approach

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 experience. 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 cognitivism, 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 successfully 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 explicitly 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.

TWO

◆ ◆ ◆

The Phenomenological Connection

THIS CHAPTER INTRODUCES a number of themes from phenomenological philosophy that will appear throughout this book. Phenomenology is important here for two main reasons. First, any attempt to gain a comprehensive understanding of the human mind must at some point consider consciousness and subjectivity—how thinking, perceiving, acting, and feeling are experienced in one's own case. Mental events do not occur in a vacuum; they are lived by someone. Phenomenology is anchored to the careful description, analysis, and interpretation of lived experience. Second, the enactive approach puts the organism and the body center-stage in mind science, but the human body, unless it is dead, is always the *lived body*. Phenomenology, in one of its strongest currents flowing from Husserl and Merleau-Ponty, is a philosophy of the lived body. For these reasons, phenomenology can guide and clarify scientific research on subjectivity and consciousness, and provide a philosophical framework for assessing the meaning and significance of this research for our self-understanding.

This chapter has two purposes. First, it introduces some core ideas of Husserl's phenomenology, in particular the phenomenological method of investigating the structure of experience, known as the phenomenological reduction, and the phenomenological concept of intentionality. Second, it sketches three phases of phenomenology, known as static, genetic, and generative phenomenology.

Static phenomenology analyzes the formal structures of consciousness, whereby consciousness is able to constitute (disclose or bring to awareness) its objects. Static phenomenology takes these intentional structures and their correlative objects as given and analyzes them statically or synchronically.

Genetic phenomenology is concerned with how these intentional structures and objects emerge through time; therefore, it cannot take them as given. Instead, it analyzes how certain types of experience motivate later and more complex types—for example, how implicit and prereflective experiences motivate attentive and reflective experiences. From the perspective of genetic phenomenology, experience has a sedimented structure, and the process of sedimentation needs to be understood in relation to the lived body and time-consciousness. Some of the key guiding phenomena for genetic phenomenology—affect, motivation, attention, habit—are familiar from the perspective of mind science, especially developmental psychology, emotion theory, and affective-cognitive neuroscience. These points of convergence and mutual illumination will be taken up in later chapters.

Whereas time-consciousness and the lived body are the guiding threads for genetic phenomenology, for generative phenomenology the guiding thread is the life-world. The subject matter of generative phenomenology is the cultural, historical, and intersubjective constitution of our human world. The importance of generative phenomenology for mind science and the enactive approach in particular will be taken up in the last chapter of this book.

Phenomenology with an Attitude

Phenomenology, in its original Husserlian inspiration, grows out of the recognition that we can adopt in our own first-person case different mental attitudes or stances toward the world, life, and experience. In everyday life we are usually straightforwardly immersed in various situations and projects, whether as specialists in scientific, technical, or practical knowledge or as colleagues, friends, and members of families and communities. Besides being directed toward these more-or-less particular, "thematic" matters, we are also directed at the world as an unthematic horizon of all our activity (Husserl 1970, p. 281). Husserl calls this attitude of being straightforwardly immersed in the world "the natural attitude," and he thinks it is characterized by a kind of unreflective "positing" of the world as something existing "out there" more or less independently of us.

In contrast, the "phenomenological attitude," arises when we step back from the natural attitude, not to deny it, but in order to investigate the very experiences it comprises. If such an investigation is to be genuinely philosophical, then it must strive to be critical and not dogmatic, and therefore it cannot take the naïve realism of the natural attitude for granted. Yet to deny this realistic attitude would be equally dogmatic. Rather, the realistic positing of the natural attitude must be suspended, neutralized, or put to one side, so that it plays no role in the investigation. In this way, we can focus on the experiences that sustain and animate the natural attitude, but in an open and nondogmatic manner. We can investigate experience in the natural attitude without being prejudiced by an unexamined view of things, which is characteristic of the natural attitude.

Yet how exactly is such an investigation to proceed? What exactly are we supposed to investigate? Husserl's answer is that our attention should be directed toward the world strictly as we experience it. We are to attend to the world strictly as it appears and as it is phenomenally manifest. Put another way, we should attend to the modes or ways in which things appear to us. We thereby attend to things strictly as correlates of our experience, and the focus of our investigation becomes the correlational structure of our subjectivity and the appearance or disclosure of the world.

The philosophical procedure by which this correlational structure is investigated is known as the *phenomenological reduction.* "Reduction" in this context does not mean replacing or eliminating one theory or model in favor of another taken to be more fundamental. It signifies rather a "leading back" *(reducere)* or redirection of thought away from its unreflective and unexamined immersion in the world to the way in which the world appears to us. To redirect our interest in this way does not mean we doubt the things before us or that we somehow try to turn away from the world to look elsewhere. Things remain before us, but we envisage them in a new way, namely, strictly as experienced. Thus, everyday things available for our perception are not doubted or considered as illusions when they are "phenomenologically reduced," but instead are envisaged and examined simply and precisely *as perceived.* Remembered things are examined strictly and precisely *as remembered,* imagined things *as imagined.* In other words, once we adopt the phenomenological attitude, we are interested not in *what* things are in some naïve, mind-independent or theory-independent sense, but rather in exactly *how* they are experienced, and thus as strict relational correlates of our subjectivity.[1]

As a procedure of working back from the what to the how of experience, the phenomenological reduction has to be performed in the first person. As is true of any such procedure, it is one thing to describe its general theoretical character and another to describe it pragmatically, the concrete steps by which it is carried out. The main methodical step crucial for the phenomenological reduction Husserl called the *epoché*. This term derives from Greek skepticism, where it means to suspend or refrain from judgment, but Husserl adopted it as a term for the "suspension," "neutralization," or "bracketing" of both our natural "positing" attitude and our theoretical beliefs and assertions (whether scientific or philosophical) about "objective reality." From a more embodied and situated, first-person perspective, however, the epoché can be described as the flexible and trainable mental skill of being able both to suspend one's inattentive immersion in experience and to turn one's attention to the manner in which something appears or is given to experience (Depraz 1999b; Depraz, Varela, and Vermersch 2000; Steinbock 2004). Suspending one's inattentive immersion in experience implies the capacity to notice such immersion, and thus implies what psychologists call meta-awareness (awareness of awareness). Being able to redirect one's attention to the manner in which something appears implies flexibility of attention; in particular it implies being able voluntarily to shift one's attention and stabilize or sustain it on a given mode of presentation. The ultimate aim is not to break the flow of experience, but to reinhabit it in a fresh way, namely, with heightened awareness and attunement.[2]

Within the phenomenological tradition one can discern a certain ambivalence regarding these theoretical and practical or existential dimensions of the epoché. On the one hand, Husserl's great concern was to establish phenomenology as a new philosophical foundation for science; thus for him the epoché served largely as a critical tool of theoretical reason.[3] On the other hand, because Husserl's theoretical project was based on a radical reappraisal of experience as the source of meaning and knowledge, it necessitated a constant return to the patient, analytic description of lived experience through phenomenological reduction. This impulse generated a huge corpus of careful phenomenological analyses of human experience—the perceptual experience of space (Husserl 1997), kinesthesis and the experience of one's own body (Husserl 1989, 1997), time-consciousness (Husserl 1991), affect (Husserl 2001), judgment (Husserl 1975), imagination and memory (Husserl 2006), and intersubjectivity (Husserl 1973), to name just a few.

Nevertheless, the epoché as a practical procedure—as a situated practice carried out in the first person by the phenomenologist—has remained strangely neglected in the phenomenological literature, even by so-called existential phenomenologists such as Heidegger and Merleau-Ponty. They instead took up and then recast in their own ways the method of the phenomenological reduction (see Heidegger 1982, pp.19–23; Merleau-Ponty 1962, pp. xi–xiv). For this reason, one new current in phenomenology aims to develop more explicitly the pragmatics of the epoché as a first-person method for investigating consciousness (Depraz 1999b; Depraz, Varela, and Vermersch, 2000, 2003; Varela and Shear 1999b). This pragmatic approach also involves comparing the epoché to first-person methods in other domains, especially Buddhist philosophy and contemplative mental training (Depraz, Varela, and Vermersch, 2003; Lutz, Dunne, and Davidson, 2007). In addition, it explores the relevance of first-person methods for producing more refined first-person reports in experimental psychology and cognitive neuroscience (Lutz and Thompson 2003). This endeavor is central to the research program known as neurophenomenology, introduced by Francisco Varela (1996) and discussed extensively later in this book.

Let us return to the phenomenological reduction in its original philosophical context. Here the reduction, in its full sense, is a rich mode of analysis, comprising two main steps. The first step leads back from the natural attitude to the phenomenological attitude by neutralizing the realistic positing of the natural attitude and then orienting attention toward the disclosure or appearance of reality to us (this step corresponds to the epoché). The second step leads from this phenomenological attitude to a more radical kind of philosophical attitude. More precisely, this step leads from phenomenology as an empirical and psychological attitude (phenomenological psychology) to phenomenology as a transcendental philosophical attitude (transcendental phenomenology).

The term *transcendental* is used here in its Kantian sense to mean an investigation concerned with the modes or ways in which objects are experienced and known, and with the *a priori* conditions for the possibility of such experience and knowledge. Husserl casts these two aspects of transcendental inquiry in a specific form that is clearly related to but nonetheless different from Kant's (see Steinbock 1995, pp. 12–15). First, transcendental phenomenology focuses not on *what things are* but on the *ways in which things are given.* For Husserl, this means focusing on phenomena (appearances) and the senses or meanings they have for us, and then asking how these meaningful phenomena are constituted (brought to awareness). Second, to address this constitutional problem, transcendental phenomenology tries to uncover the essential formal laws under which experience necessarily operates in order to constitute a meaningful world.

In the natural attitude, reality is taken for granted as being simply there without any active engagement on the part of consciousness. In other words, there is no thought that reality involves acts or processes of constitution. Grasped phenomenologically, in the transcendental phenomenological attitude, reality is that which is disclosed to us as real, whether in everyday perception or scientific investigation, and such disclosure is an achievement of consciousness. The point here is not that the world would not exist if not for consciousness. Rather, it is that we have no grip on what reality means apart from what is disclosed to us as real, and such disclosure necessarily involves the intentional activity of consciousness. The point of the transcendental phenomenological reduction is to gain access to this activity and the constitutional role it plays.

It is often said that whereas Husserl's orientation is transcendental in this way, Heidegger and Merleau-Ponty reject the transcendental standpoint and identify the constitutional structures unearthed by phenomenology with existential structures of "being-in-the-world." (The hyphens indicate that "being," "in," and "world" are not ontologically separable, but form one irreducible and unified structure.) But this interpretation is simplistic. First, both Heidegger's "Dasein" (his term for individual human existence) and Merleau-Ponty's "lived body" (a concept that comes straight from Husserl) are transcendental in the relevant sense, for they are ways of characterizing that which makes possible the disclosure or manifestation of the world as meaningful.

Second, although Husserl in the most well-known portions of his work (the writings published in his lifetime) did focus largely on the constitutional structures of "egological" consciousness (consciousness at the level of the individual reflective "I" or "ego"), recent scholarship indicates that these analyses are not fully representative of his mature philosophical investigations.[4] As his thought developed, he greatly expanded his investigations, analyzing constitutional structures belonging to the "nonegological" (or "pre-egological") depths of the lived body, time-consciousness, and intersubjectivity, as well as the terrain of historical and cultural life.[5] The point here is more than an interpretive or textual one; it is philosophical. Transcendental phenomenology cannot be limited to—and indeed goes far beyond—a philosophy of "egological" consciousness or subjectivity. "Transcendental" signifies a radical attitude, one that aims to regress back to the very roots (conditions of possibility) of our experience of a meaningful world. These roots ramify far beyond individual consciousness into the depths of our lived bodies and out into our social and cultural worlds.

The remainder of this chapter sketches a few of these developments of phenomenological thought. My aim is not to give a detailed scholarly account of any particular aspect of phenomenology, but to set forth some themes and ideas important for the chapters to come.

Intentionality

A good place to begin is the phenomenological doctrine of the intentionality of consciousness. According to phenomenology, consciousness is intentional, in the sense that it "aims toward" or "intends" something beyond itself. This sense of intentional should not be confused with the more familiar sense of having a purpose in mind when one acts, which is only one kind of intentionality in the phenomenological sense. Rather, *intentionality* is a generic term for the pointing-beyond-itself proper to consciousness. (It comes from the Latin *intendere,* which once referred to drawing a bow and aiming at a target.)

Phenomenologists distinguish different types of intentionality. In a narrow sense, they define intentionality as object-directedness. In a broader sense, they define it as openness to the world or what is "other" ("alterity"). In either case, the emphasis is on denying that consciousness is self-enclosed.[6]

Object-directed experiences are those in which we are conscious *of* something in a more-or-less determinate sense. When we see, we see something; when we remember, we remember something; when we hope or fear, we hope for or fear something. These kinds of "transitive consciousness" are characterized by the intending of an object (which need not exist). "Object" in its etymological sense means something that stands before us. Something standing before us lies beyond, over against, or outside of us. Object-directed experiences can thus be understood as experiences in which we are conscious of something distinct from ourselves as a present subject, whether it be a past event remembered, something perceived in the settings around us, a future event feared or hoped for, something imagined, and so on.

Many kinds of everyday experience, however, are not object-directed in this sense. Such experiences include bodily feelings of pain, moods such as undirected anxiety, depression, and elation, and absorbed skillful activity in everyday life. These experiences are not or need not be "about" any intentional object. They are not directed toward a transcendent object, in the sense of something experienced as standing over against oneself as a distinct subject. Put another way, they do not have a clear subject-object structure.[7]

Philosophers who think of intentionality simply as object-directedness would deny that experiences like these are intentional. Nevertheless, such experiences do qualify as intentional in the broader phenomenological sense of being open to what is other or having a world-involving character. Thus bodily feelings are not self-enclosed without openness to the world. On the contrary, they present things in a certain affective light or atmosphere and thereby deeply influence how we perceive and respond to things. A classic example is Sartre's discussion of feeling eyestrain and fatigue as a result of reading late into the night (1956, pp. 332–333).[8] The feeling first manifests itself not as an intentional object of transitive consciousness but as a trembling of the eyes and a blurriness of the words on the page. One's body and immediate environment disclose themselves in a certain manner through this feeling. In the case of moods, although they are not object-directed in the same way intentional emotions are—such as a feeling of sympathy for a loved one or a feeling of envy for a rival—they are nonetheless hardly self-enclosed without reference to the world. On the contrary, as Heidegger analyzes at length in *Being and Time,* moods reveal our embeddedness in the world and (as he sees it) make possible more circumscribed forms of directedness in everyday life. Finally, in absorbed skillful activities, such as driving, dancing, or writing, one's experience is not that of relating to a distinct intentional object but of being engaged and immersed in a fluid activity. Such experience takes on a subject-object structure only during moments of breakdown or disruption (see Dreyfus 1991, 2002; Dreyfus and Dreyfus 1986).

In phenomenology, intentional experiences are described as mental *acts*—acts of perceiving, remembering, imagining, empathizing, and so on. Phenomenology conceives of mental life as a temporally extended and dynamic process of flowing intentional acts. These acts are animated by precognitive habits and sensibilities of the lived body. Intentional acts are performances of a person, a living bodily subject of experience, whose cognitive and affective life is constituted by communal norms, conventions, and historical traditions. Mental life is animated by an intentional striving that aims toward and finds satisfaction in disclosure of the intentional object. In this way, intentionality is teleological (Held 2003, p. 14).

Given this conception of intentionality, it follows that neither the mental act nor that which it intends can be understood in isolation. Every mental act is the very act it is in virtue of that which it intends, and every object is constituted in and through the temporally extended course of intentional experience. As Donn Welton explains:

> There is a genuinely new conception of mental acts here in play . . . On the one hand, acts do not belong to a closed interior realm available only to introspection. Rather, they have their being by virtue of their relationship to that which transcends them. On the other hand, the determinations of "the given" can be fully clarified only by seeing them in relation to certain acts that contribute to their configuration. It is neither the subject nor the object but the relationship that is primary. (Welton 2000, p. 17)

Phenomenologists call this relation the correlational structure of intentionality. "Correlational" does not mean the constant conjunction of two terms that could be imagined to exist apart; rather, it refers to the invariant structure of intentional act/intentional object. Object-directed intentional experiences necessarily comprise these two inseparable poles. In Husserlian phenomenological language, these two poles are known as the "noema" (the object in its givenness) and the "noesis" (the mental act that intends and discloses the object in a certain manner).[9]

We need to keep this framework in mind when we think about the relation between the phenomenological conception of intentionality and what philosophers of mind today call *mental representation.* In a broad and theoretically neutral sense, a mental representation is supposed to be a mental structure (concept, thought, image) with semantic properties (content, truth conditions, reference), or a state or process involving such a structure. Usually, a mental representation is not considered to be an object of cognition or awareness, but rather that by which one cognizes or is aware of something in the world. Many phenomenologists would agree that intentional experience is representational in this broad sense of having descriptive content—that in intentional experience the world is represented in some particular way or other. Nevertheless, the phenomenological conception of intentional experience has certain other distinctive features. First, in phenomenology, as mentioned earlier, intentional experiences are conceptualized not as *states having content* but as *acts having directedness.* These two conceptions are not necessarily incompatible, but their theoretical orientation and emphasis are different. Second, "re-presentation," in its technical phenomenological sense, applies only to certain types of intentional acts, namely, those that mentally evoke or bring to presence something that is not present in its bodily being.

Phenomenologists thus draw a crucial distinction between intentional acts of *presentation (Gegenwärtigung)* and of *re-presentation (Vergegenwärtigung)* (see Marbach 1993). On the one hand, perceptual experience is presentational: in this type of experience the object is given as present in its very being. In memory or imagination, on the other hand, the object imagined or remembered is not given as present in its very being, but rather as both phenomenally absent and as mentally evoked or called forth. In this way, memory and imagination are said to be re-presentational. Note that the definitive feature of re-presentational experience is that the object is given as absent and as mentally evoked, but not necessarily as *re*-evoked or called forth *again.* Re-evoking belongs to memory but not necessarily to visualizing or fantasizing. Note also that re-presentational experiences do not float freely, as it were, but arise in relation to ongoing presentational experiences of one's surroundings. I discuss this phenomenological conception of mental re-presentation in connection with mental imagery in Chapter 10.

Let us return to the connection between phenomenology and the enactive approach. As we will see in Chapter 3, the main explanatory tool of the enactive approach is the theory of self-organizing and autonomous dynamic systems. Such systems bring forth or enact meaning in continuous reciprocal interaction with their environments. "Inner" and "outer" are not preexisting separate spheres, but mutually specifying domains enacted or brought forth by the structural coupling of the system and its environment. This subpersonal account of cognitive systems echoes the personal-level account of the correlational structure of intentionality.[10] As Jean-Pierre Dupuy writes in his philosophical history of cognitive science, discussing the "missed encounter" between phenomenology and mind science in the cybernetic era:

> A given [autonomous] network usually possesses a multiplicity of self-behaviors (or, as they are sometimes called, "attractors" . . .) and converges toward one or another of them depending on the initial conditions of the network. The "life" of a network can thus be conceived as a trajectory through a "landscape" of attractors, passing from one to another as a result of perturbations or shocks from the external world. Note that these external events come to acquire meaning in the context of the network as a result of the network's own activity: the *content*—the meaning—that the network attributes to them is precisely the self-behavior, or attractor that results from them. Obviously, then, this content is purely endogenous and not the reflection of some external "transcendent" objectivity.
>
> It should be obvious, too, that this line of argument . . . provides us with at least the germ of a very satisfactory model of what Brentano called "immanent objectivity" . . . The attractor is an entity that both fully participates in the activity of the network and yet in some sense, by virtue of the fact that it results from a higher level of logical complexity, transcends the activity of the network. The dynamics of the network may therefore be said to *tend toward* an attractor, although the latter is only a product of these dynamics. The network is thus an *intentional* creature in Brentano and Husserl's sense. Systems theory was to coin

> another term to describe this paradoxical relationship between the dynamics of a system and its attractor, referring to it as "autotranscendence." This is not very different, really, from Husserl's notion of "transcendence within immanence." (Dupuy 2000, pp. 104–105)

Because this notion of transcendence-within-immanence is often misunderstood, it bears clarification here. It does not mean that what appears to be beyond or outside the sphere of mental activity is really contained within the mind (in some idealist or internalist sense). Rather, the crucial point is that the transcendent is given as such by virtue of the intentional activities of consciousness. Thus it falls within the sphere of what is phenomenologically constituted (disclosed or brought to awareness by consciousness). Clearly, this point makes sense only at a transcendental level, for at this level the transcendent is understood as a *mode of givenness* or disclosure (one characterizing things in the world, but not one's own consciousness). Thus, at a transcendental level, what is *really or genuinely transcendent* is also *phenomenologically immanent* (see Crowell, in press, for further discussion).

The correspondence between phenomenology and dynamic systems theory to which Dupuy is pointing should therefore be understood as follows. External events are *really transcendent,* for they are certainly not contained within the system, nor are they a mere product of what goes inside the system. Nevertheless, they are *intentionally immanent,* in the following sense: they do not arrive already labeled, as it were, as external events; instead they are constituted or disclosed as such, and with the significance they have, by virtue of the network's autonomous (self-organizing) dynamics. In other words, their status as external events for the system (as opposed to their status for an observer of the system) is a function of the system's own activity. Their meaning or significance corresponds to an attractor of the system's dynamics (a recurrent pattern of activity toward which the system tends), which itself is an emergent product of that very dynamics. The external world is constituted as such for the system by virtue of the system's self-organizing activity. Dupuy's proposal, in a nutshell, is that *constitutional intentionality corresponds to a kind of self-organization.* This proposal, as we will see in later chapters, is one of the key guiding intuitions of the enactive approach and neurophenomenology.

From Static to Genetic Phenomenology

This convergence between phenomenology and the enactive approach can be taken further. The correlational structure of intentionality belongs to what Husserl called static phenomenology. As his thought progressed, however, Husserl found that he needed to articulate a *genetic phenomenology,* that is, a phenomenology whose point of departure is not the explicit correlational structure of intentional act (noesis) and intentional object (noema), but rather the genesis of intentional experience in time. From the standpoint of genetic phenomenology, we need to account for the correlational structure of intentionality developmentally by understanding how it emerges from inarticulate experience that does not have a clear subject-object structure. One wellspring of this kind of experience is the lived body *(Leib);* another is time-consciousness. The shift from static to genetic phenomenology thus marks a turn toward the lived body and time-consciousness. Thus it enables us to deepen the connection between phenomenology and the enactive approach.

Static phenomenology makes use of two methodological strategies (Steinbock 1995, pp. 38–39). The first is static analysis or the analysis of invariant formal structures of experience, such as the correlational structure of intentionality, or the difference between presentational and re-presentational mental acts and the ways the latter presuppose the former. The second strategy is constitutional analysis—the analysis of how things are disclosed or brought to awareness by virtue of the intentional activities of consciousness. From a transcendental perspective, the invariant formal structures of experience uncovered by static analysis are precisely the essential formal laws under which experience necessarily operates in order to constitute its objects. An example is Husserl's investigation in his 1907 lectures, "Thing and Space," of the conditions of possibility for the perceptual experience of things in space (Husserl 1997). Husserl shows that visual perception depends constitutively on certain invariant functional interdependencies between visual sensation and the experience of moving one's body (which he calls kinesthesis). These analyses anticipate recent enactive or dynamic sensorimotor accounts of perception (discussed in Chapter 9). According to these accounts, to perceive is to exercise one's skillful mastery of the ways sensory stimulation varies as a result of bodily movement (Noë 2004; O'Regan and Noë 2001a).

Unlike static phenomenology, genetic phenomenology does not take the already disclosed intentional object as its point of departure, nor is it content to stay at the level of analyzing formal and constitutive structures of experience. Instead, it investigates the genesis and development of those structures themselves. After all, we do not simply drop into the world and open our eyes and see. What we see is a function of how we see, and how we see is a function of previous experience. For genetic phenomenology, what we experience is not a fixed given but something that has come to be given—something *emergent*—out of previous experience (Bernet, Kern, and Marbach 1993, pp. 200–201). In Chapters 11 and 12, I discuss phenomenological analyses of time-consciousness and affect produced from this genetic orientation and relate them to research in psychology and neuroscience.

Genetic phenomenology also brings with it a different way of thinking about the conscious subject. From a static viewpoint, the "I" is thought of as a kind of "ego-pole" of the noetic-noematic structure, in contraposition to the "object-pole."[11] A fuller articulation of the correlational structure of intentionality would thus be [ego] noesis-noema (I intend the intentional object). From a genetic standpoint, however, this way of thinking remains abstract because it ignores the temporal development and individuation of the subject. The "I" or "ego" is not a mere "empty pole" of selfhood in experience but a concrete subject having habits, interests, convictions, and capabilities as a result of accumulated experience. In other words, the subject has to be seen as having a "life" in all the rich senses of this word—as formed by its individual history, as a living bodily subject of experience *(Leib),* and as belonging to an intersubjective "life-world" *(Lebenswelt).*

Genetic phenomenology distinguishes between active genesis and passive genesis. In active genesis subjects play an active and deliberate, productive role in the constitution of objects. The products of active genesis are tools, artworks, scientific theories, experimental interventions, logical judgments, mathematical propositions, and so on. Active genesis, however, always presupposes a passivity by which one is *affected* beforehand. It must be stressed that "passive" in this context does not mean a state of inactivity, but rather a state of being involuntarily influenced and affected by something. In particular, it means being influenced and affected on an aesthetic level, in the original Greek sense of *aisthesis* as sense perception, including especially the perception and felt experience of what is attractive and unattractive. Thus the thought behind the active / passive distinction is that our active orientation toward things in practical or theoretical reason, or artistic creation, presupposes a deeper and more fundamental openness to the world. It is an openness to being sensuously affected and solicited by the world through the medium of our living body, and responding by attraction and repulsion. Investigating these sensorimotor and affective depths of experience leads phenomenology to the notion of passive genesis. In passive genesis, the lived body constitutes itself and its surrounding environment through the involuntary formation of habits, motor patterns, associations, dispositions, motivations, emotions, and memories.

At this level of "passive synthesis" in experience, the relevant notion of intentionality is not so much object-directedness as openness to the world, here in the bodily form of an implicit sensibility or sentience that does not have any clear subject-object structure. Intentionality at this level functions anonymously, involuntarily, spontaneously, and receptively. Husserl distinguishes between receptivity and affectivity (2001, pp. 105, 127). As Dan Zahavi explains, "Receptivity is taken to be the first, lowest, and most primitive type of intentional activity, and consists in responding to or paying attention to that which is affecting us passively. Thus, even receptivity understood as a mere 'I notice' presupposes a prior affection" (Zahavi 1999, p. 116; see also the Translator's Introduction to Husserl 2001). Affection here means being *affectively influenced* or *perturbed*. The idea is that whatever comes into relief in experience must have already been affecting us and must have some kind of "affective force" or "affective allure" in relation to our attention and motivations. Whatever exercises affective allure without our turning to it attentively is said to be "pregiven," and whatever succeeds in gaining attention is said to be "given." Thus the given—the mode or way in which something appears to object-directed consciousness—has to be understood dynamically and teleologically as emergent in relation to the pregiven. Object-directed intentional experiences emerge out of the background of a precognitive "operative intentionality" (Merleau-Ponty 1962, p. xviii) that involves a dynamic interplay of affective sensibility, motivation, and attention. This affectively "saturated intentionality" (Steinbock 1999) provides our primordial openness to the world.

The phenomenological terrain of "passive synthesis" is rich in potential for illuminating and being illuminated by research in psychology and neuroscience on emotion and cognition. Some of these connections are already discernible in Husserl's description of passive synthesis as operating according to a principle of association (Husserl 1960, pp. 80–81; 2001, pp. 162–242). For Husserl, association is an intentional process whereby experiences are built up or synthesized into larger, patterned wholes. Using the terminology of emergence, we could say that association is the process by which coherent patterns of experience emerge from conjoined and reciprocally affecting experiences. Here are a few vivid examples given by William James in his *Principles of Psychology:*

> Let a person enter his room in the dark and grope among the objects there. The touch of the matches will instantaneously recall their appearance. If his hand comes in contact with an orange on the table, the golden yellow of the fruit, its savor and perfume will forthwith shoot through his mind. In passing the hand over the sideboard or in jogging the coal-scuttle with the foot, the large glossy dark shape of the one and the irregular blackness of the other awaken in a flash and constitute what we call the recognition of objects. The voice of the violin faintly echoes through the mind as the hand is laid upon it in the dark, and the feeling of the garments or draperies which may hang about the room is not understood till the look correlative to the feeling has in each case been resuscitated . . . But the most notorious and important case of the mental combination of auditory with optical impressions originally experienced together is furnished by language. The child is offered a new and delicious fruit and is at the same time told that it is called a "fig." Or looking out of the window he exclaims, "What a funny horse!" and is told that it is a "piebald" horse. When learning his letters, the sound of each is repeated to him whilst its shape is before his eye. Thenceforward, long as he may live, he will never see a fig, a piebald horse, or a letter of the alphabet without the name which he first heard in conjunction with each clinging to it in his mind; and inversely he will never hear the name without the faint arousal of the image of the object. (1981, pp. 524–525)

According to the empiricist philosophers Locke and Hume, such associations happen in a completely mechanical way. Association operates as a kind of connective force in the mind that links impressions and ideas simply in virtue of their simultaneous occurrence, proximity, or repeated succession. Hume's analysis of causation provides a famous example of this way of thinking about association. Hume argued that causal connections are neither directly observable nor provable by reason, but are objects of mere belief based on habit and custom. The belief in a causal connection between A and B arises from the association or “constant conjunction” of A and B in past experience: experiences of A constantly conjoined to experiences of B make the mind habitually expect that the occurrence of A will be followed by the occurrence of B.

For Husserl (and James), however, association is not meaningless and mechanical, but thoroughly intentional. Association is not the mechanical aggregation of complex experiences out of preexisting experience-atoms. Husserl, like James, completely rejects this atomistic conception of experience. Like emergent processes in a self-organizing system, associated experiences reciprocally strengthen and reinforce each other and thereby give rise to new formations that supersede their prior separateness. Association also involves the retention and anticipation of sense or meaning. Earlier experiences are affectively “awakened” by later ones on the basis of their felt similarities, and they motivate the anticipation that what is to come will cohere with the sense or meaning of experience so far. In Husserl’s terminology, there is an “analogical transfer of sense” from earlier to later experience: what is present now is passively apprehended within a sense that has its roots in earlier experience and that has since become habitual (Bernet, Kern, and Marbach 1993, p. 202).

The notion of habit is central to Husserl’s conception of passive genesis, as he states explicitly in a lecture from 1927: “As Hume correctly teaches, habit is not only our nurse, rather it is the function of consciousness that shapes and constantly further shapes the world” (quoted by Bernet, Kern, and Marbach 1993, p. 203; see also Welton 2000, p. 243). Husserl mentions Hume, but the notion of habit was very important to James as well. In his *Principles of Psychology* James declared that habit is the ground of all association in the stream of consciousness and in brain activity (thereby anticipating Donald Hebb and connectionism).[12] Later, in 1945, Merleau-Ponty introduced his notion of the habit-body in his *Phenomenology of Perception* while discussing the experience of the phantom limb: “our body comprises as it were two distinct layers, that of the habit-body and that of the body at this moment. In the first appear manipulatory movements which have disappeared from the second, and the problem how I can have the sensation of still possessing a limb which I no longer have amounts to finding out how the habitual body can act as guarantee for the body at this moment" (1962, p. 82). To say that the habitual body acts as guarantee for the body at this moment is to say that one's lived body is a developmental being thick with its own history and sedimented ways of feeling, perceiving, acting, and imagining. These sedimented patterns are not limited to the space enclosed by the body's membrane; they span and interweave the lived body and its environment, thereby forming a unitary circuit of lived-body-environment (Gallagher 1986b).

In Part III, I will explore this convergence of genetic phenomenology and enactive cognitive science in greater detail. For now let me simply point out how important the dynamic coupling between one's lived body and the surrounding world is to both perspectives. In this coupling, the motivating undercurrent is the habitual and associative linkage of affective, sensorimotor, and imaginative bodily experiences.

From Genetic to Generative Phenomenology

Late in his life Husserl began to move in still another direction—from genetic phenomenology to *generative phenomenology*. Already in genetic phenomenology intersubjectivity had arisen as an important theme, in the form of the dynamic coupling between self and other on the basis of their lived bodily presence to one another. Generative phenomenology, however, widened the scope of this genetic analysis beyond the self-other relation to include the parameters of birth and death as well as the interconnectedness of generations.

In this context, the term *generative* has a double meaning: it means both the process of becoming and the process of occurring over the generations (Steinbock 1995, p. 3). Generative phenomenology concerns the historical, social, and cultural becoming of human experience. If static phenomenology is restricted in scope with respect to genetic phenomenology, then genetic phenomenology is restricted in scope with respect to generative phenomenology: the subject matter of generative phenomenology is the historical and intersubjective becoming of human experience, whereas genetic phenomenology focuses on individual development without explicit analysis of its generational and historical embeddedness.

In shifting from a genetic to a generative register, the notion of the lived body is complemented with that of the *life-world* (Husserl 1970; Steinbock 1995, pp. 86–122; Welton 2000, pp. 331–346). The life-world is the everyday world in which we live. It is "always already pregiven," serving as the horizon of all our activities, practical and theoretical. Two important aspects of this rich and multifaceted notion need to be mentioned here—the back-and-forth circulation or exchange within the life-world between empirical science and everyday human life, and the life-world as the pregiven horizon and ground of all human activity.

The life-world comprises the everyday world and the things that can be directly experienced within the everyday world—our living bodies, our natural surroundings, and our cultural creations (tools, artworks, and so on). The life-world is subject-relative in the sense that it is relationally bound to human subjectivity. This is in contrast to "objective nature" as conceived by science, which is arrived at through logical and theoretical abstraction. *Nature* so construed is an objectification and has as its cognitive correlate the objectifying intentional attitude adopted by a community of theorizing subjects. *Objective nature* presupposes the life-world as its evidential source and ground. In principle it cannot be experienced directly because it is the product of abstraction and idealization. Nevertheless, the propositions, models, logical constructs, and experimental techniques of the sciences are clearly experienceable in another sense: they are human accomplishments that have experiential validity for members of the scientific community, and their effects flow into the everyday world and become tangibly experienced in the form of technology and social practice. Our life-world encompasses science, in addition to other spheres of experience such as art, philosophy, and religion. Hence, there is a necessary "circulation" between everyday experience and scientific experience (Varela, Thompson, and Rosch 1991, pp. 10–14). On the one hand, everyday experience provides the sensuous, material contents from which and with which science must work. On the other hand, the scientific analyses built from these contents contribute to the formation of our life-world and provide important leading clues for phenomenological analyses of how our experience of the world is genetically and generatively constituted.

In taking up these phenomenological analyses, Husserl initially conceived of the life-world as a synthetic totality. Hence he treated it on the model of an object, albeit a peculiar all-encompassing one (see Steinbock 1995, pp. 98–102; Welton 2000, pp. 336–346). Eventually, however, it became clear to him that the life-world cannot be given as any kind of intentional object, for it is always already there, pregiven rather than given.[13] Thus, in a crucial and famous passage from his last work, *The Crisis of European Sciences and Transcendental Phenomenology* (1970, §37, pp. 142–143), he wrote that the world is always already there, existing in advance for us, as the "ground" and "horizon" of any human activity. He then asserted that the way we are conscious of the world and the way we are conscious of things or objects, though inseparably united, are fundamentally different. We can be conscious of things only as things within the world horizon, and we can be conscious of the world horizon only as a horizon for existing objects. Yet the world is not any kind of entity, nor is it simply the totality of entities, precisely because it is the horizon presupposed by any entity or any totality. It is tempting to say "the world is one," except that, as Husserl puts it, the world "exists with such uniqueness that the plural makes no sense when applied to it." In other words, the world is not one in any sense in which it could have been two. To put it another way, to describe the world as "unique," such that "every singular and every plural drawn from it, presupposes the world horizon," means that the notion of counting makes no sense or has no application here.[14] Given this difference between the manner in which any object is given and the manner in which the world horizon is given (namely, as always already pregiven), it follows that there must be "fundamentally different correlative types of consciousness for them."

Husserl's terms *horizon* and *ground* are metaphorical, at once visual and geological. A horizon is not a thing "out there" but rather a structure of appearance. It therefore implicates or points back to the perceiver for whom appearances are so structured. In phenomenological language, "horizon" taken noematically as a structure of appearance necessarily implicates "horizon" taken noetically as a structure of consciousness. One could say that a horizon is the precondition for the appearance of anything, except that "precondition" is too static. Stated in a genetic register, a horizon is a dynamic structure of disclosure in which both the object (noema) and consciousness (noesis) partake (Steinbock 1995, p. 107). Anything that comes forth, manifests, or emerges does so in an open clearing or expanse, delimited by a horizon. The horizon of every possible horizon is the world. Yet the world-horizon cannot be the synthesis, totality, or mereological sum of all these possible horizons because it is pregiven or *a priori* with respect to any of them and thus is *sui generis.* Similarly, to describe the life-world as ground *(Boden)* is not to say that it is a static foundation; rather, it is the pregiven soil out of which everything is generated and nourished. This soil includes one's forebears and culture. We human beings constitute and reconstitute ourselves through cultural traditions, which we experience as our own development in a historical time that spans the generations. To investigate the life-world as horizon and ground of all experience therefore requires investigating none other than generativity—the processes of becoming, of making and remaking, that occur over the generations and within which any individual genesis is always already situated.

Generative phenomenology brings to the fore the intersubjective, social, and cultural aspects of our radical embodiment. Individuals are born and die, they develop and constantly change, and they emerge from their forebears and perpetuate themselves in generations to come. Individual subjectivity is from the outset intersubjectivity, originally engaged with and altered by others in specific geological and cultural environments (Depraz 1999c, p. 482; Steinbock 1995). Individual subjectivity is intersubjectively and culturally embodied, embedded, and emergent.

Classical cognitive science, to the extent that it operated under the assumption that the individual self comes first and the other second, simply left out intersubjectivity and culture. Indeed, it had no real means to analyze their contributions to the "cognitive architecture" of the human mind. As a result, classical cognitive science has offered abstract and reified models of the mind as a disembodied and cultureless physical symbol system or connectionist neural network in the head of a solitary individual. As we will see in the last chapter of this book, however, the enactive approach, particularly when guided by genetic and generative phenomenologies of the lived body, intersubjectivity, and the life-world, offers a different vision. I will argue that self and other enact each other reciprocally through empathy and that human subjectivity emerges from developmental processes of enculturation and is configured by the distributed cognitive web of symbolic culture.

THREE

◆ ◆ ◆

Autonomy and Emergence

ACCORDING TO THE ENACTIVE APPROACH, the human mind emerges from self-organizing processes that tightly interconnect the brain, body, and environment at multiple levels. The key ideas on which this proposition is based are those of *autonomous systems* and *emergence* or *emergent processes*. In this chapter, I explain these ideas. In the next chapter, I explore some connections between these ideas and phenomenological ideas about form, in particular forms or structures of behavior. These two chapters will lay the groundwork for the enactive strategy of addressing the explanatory gap by going back to the roots of mind in life and then working forward to subjectivity and consciousness.

In the first section of this chapter, I review some basic ideas about dynamic systems that form a background for the enactive approach. In the second section, I explain the notion of an autonomous system. A distinctive feature of the enactive approach is the emphasis it gives to autonomy. In brief, an autonomous system is a self-determining system, as distinguished from a system determined from the outside, or a heteronomous system. On the one hand, a living cell, a multicellular animal, an ant colony, or a human being behaves as a coherent, self-determining unity in its interactions with its environment. An automatic bank machine, on the other hand, is determined and controlled from the outside, in the realm of human design. The paradigm for interaction with a heteronomous system is input/processing/output, in which deviations from desired outputs are seen as system errors. The paradigm for interaction with an autonomous system is a conversation, in which unsatisfactory outcomes are seen as breaches of understanding (Varela 1979, p. xii). According to the enactive approach, living beings and cognitive agents need to be understood as autonomous systems. I discuss the implications of this autonomy perspective for how we think about information in the third section.

In the fourth section, I turn to emergence, a now familiar notion that describes the arising of large-scale, collective patterns of behavior in complex systems as diverse as cells, brains, ecosystems, cities, and economies. Emergence is closely related to self-organization and circular causality, both of which involve the reciprocal influence of "bottom-up" and "top-down" processes. For example, a tornado emerges through the self-organization of circulating air and water particles; it reciprocally sucks those particles into a particular macroscopic configuration, with devastating effect for anything in its path. In this section, I sketch a way of thinking about emergence that I call dynamic co-emergence. Dynamic co-emergence means that a whole not only arises from its parts, but the parts also arise from the whole. Part and whole co-emerge and mutually specify each other. A whole cannot be reduced to its parts, for the parts cannot be characterized independently of the whole; conversely, the parts cannot be reduced to the whole, for the whole cannot be characterized independently of the parts. I discuss philosophical issues related to this conception of emergence in Appendix B.

Dynamic Systems

In recent years growing interest in the dynamics of cognition and emotion has given rise to a distinct dynamical approach in mind science (Kelso 1995; Lewis and Granic 2000; Port and van Gelder 1995; Thelen and Smith 1994). The central idea of the dynamical approach is that natural cognition—cognition in evolved, living agents—is a dynamic phenomenon and accordingly needs to be understood from the perspective of the science of dynamic systems. This perspective includes dynamic-systems theory (a branch of pure mathematics), dynamic-systems modeling (mathematical modeling of empirical systems), and experimental investigations of biological and psychological phenomena informed by these tools.

The first important concept we need to introduce in this context is that of a *dynamic system.*[1] In simple terms a dynamic system is one that changes over time. The term *system*, however, is ambiguous, in that it can refer either to an actual system in the world, such as the solar system, or to a mathematical model of an actual system. In the case of the actual world, the term *system* does not admit of precise definition. In general, a system is a collection of related entities or processes that stands out from a background as a single whole, as some observer sees and conceptualizes things. The classic example from the history of science is the solar system. Its components are the sun, moon, and planets, and its states are their possible configurations. What changes over time is the state of the system. A dynamic system in the sense of a model, however, is a mathematical construction that aims to describe and predict the way an actual system changes over time (the paths of the planets, and events such as eclipses, in the case of the solar system). To this end, some aspects of the actual system are singled out as being especially important and are mathematically represented by quantitative variables. Specifying the numerical values of all the variables at a given time indicates the state of the system at that time. A dynamic system includes a procedure for producing such a description of the state of the system and a rule for transforming the current state-description into another state-description for some future time. A dynamic system is thus a mathematical model for the way that a system changes or behaves as time passes.

If the passage of time is considered to be continuous (like the sweeping second hand of an analogue clock), then the dynamic system is a differentiable one: the variables change in a smooth and continuous way, and the rules or "evolution equations" that govern the changing state of the system take the form of differential equations. If time is considered to pass in evenly spaced, discrete jumps (like a digital clock), then the system is described by a difference equation or a mapping (a function repeatedly applied or iterated in discrete time steps). Some differential equations have an analytical solution, which means they can be exactly solved by mathematical formulas. Given the starting values of the variables (the initial conditions), then all future states of the system can be known without recalculating the state of the system for each time increment. Most differential equations, however, cannot be solved in this way. When the equations contain nonlinear terms—functions in which the value of the output is not directly proportional to the sum of the inputs—then such a solution is impossible.

Therefore a different mathematical approach has to be taken from that of finding a formula that makes possible the prediction of a future state from a present one.

This other approach, introduced by Henri Poincaré in the nineteenth century, is known as the *qualitative* study of differential equations (or of nonlinear differentiable dynamic systems). One thinks of the space of all possible states of the system as a geometric space, known as state space or phase space, and the way that the system changes or behaves over time as curves or trajectories in this space. Instead of seeking a formula for each solution as a function of time, one studies the collection of all solutions (corresponding to trajectories in phase space) for all times and initial conditions at once (Norton 1995, p. 46). This approach is said to be qualitative because it uses topological and geometrical techniques to study the general or global character of the system's long-term behavior (its behavior in phase space), instead of seeking to predict the system's exact future state (the specific values of its variables at a future time). It is precisely this qualitative approach to dynamics that goes by the name of dynamic systems theory.

We need to introduce one more related notion—that of complexity. The term *complexity* describes behavior that is neither random nor ordered and predictable; rather, it is in between, exhibiting changing and unstable patterns. Of particular importance in the context of recent nonlinear dynamic-systems approaches to the brain and behavior is the notion of *complexity as dynamic instability* or *metastability*—"the successive expression of different transient dynamics with stereotyped temporal patterns being continuously created and destroyed and reemerging again" (Friston 2000b, p. 238). Recent science indicates that complexity of this sort can be found at numerous scales and levels, from the molecular and organismic to the ecological and evolutionary, as well as the neural and behavioral.[2] In every case the message seems to be that complexity, instability, or metastability is necessary for self-organization and adaptive behavior.

We can now return to the dynamical approach in mind science. The fundamental dynamical hypothesis of this approach is that natural cognitive agents (people and other animals) are dynamic systems (or, more precisely, that the cognitive systems agents instantiate are dynamic systems), and that accordingly action, perception, and cognition should be explained in dynamic terms (van Gelder 1998). Proponents of the dynamical hypothesis contrast it with the cognitivist hypothesis, which states that cognitive agents (or the cognitive systems they instantiate), whether natural or artificial, are digital computers or physical symbol systems and that accordingly cognition should be explained in symbol-processing terms.

To illustrate these ideas, we can turn to research on neural and behavioral coordination dynamics by Haken, Kelso, and colleagues (Bressler and Kelso 2001; Kelso 1995). One case they have studied is rhythmic finger movement (Haken, Kelso, and Bunz 1985). The experimental task was to move the two index fingers at the same frequency from side to side. At low speeds, there are two comfortable coordination patterns (the system is bistable): either the fingers move in-phase (equivalent muscle groups in each hand contract simultaneously) or anti-phase (equivalent muscle groups alternate in their contraction and expansion). As the speed gradually increases, the in-phase pattern becomes unstable, and eventually at a certain critical frequency the fingers spontaneously switch to an anti-phase pattern (the system undergoes a bifurcation). As the speed decreases, the in-phase pattern becomes stable again, but it does so below the original switching point (this delayed return to a previous state is known as hysteresis).

Haken, Kelso, and colleagues devised a dynamic-systems model to describe and predict these properties of motor behavior. The model describes how the relative phase relation between the two fingers evolves over time. Relative phase is an example of a "collective variable"—one whose value is set by the relation between the values of other variables, in this case those describing the individual finger movements. A collective variable describes a high-level or global characteristic of a system that emerges as a coherent and ordered pattern from the interactions of the system's components. This macrolevel pattern is also known as an order parameter because it reduces the degrees of freedom of the system's components by organizing them into a coherent and ordered pattern. When the fingers move in-phase, the collective variable or order parameter of relative phase is zero; once the critical transition or bifurcation to anti-phase happens, the relative phase becomes nonzero up to some maximum value. Because the phase transition occurs at a certain critical frequency of finger oscillation, the frequency acts as a “control parameter” for the system. The control parameter does not dictate or prescribe the collective variable or order parameter (the emergent pattern of relative phase). Rather, its changing values lead the system through a variety of possible patterns or states (Kelso 1995, p. 7). Thus the model mathematically describes how the control parameter of finger-movement frequency leads the system through different patterns of finger coordination.

In the language of dynamic-systems theory, this kind of description gives the state space of the system—the abstract and multidimensional space that represents all possible states of the system by specifying all possible values of the system’s variables. The temporal evolution of the system corresponds to its trajectory through this space. The model predicts the observed switching from one phase to another without positing any internal motor program that directs the switches by issuing symbolic instructions. Instead, the phase transitions occur spontaneously as emergent properties of the system’s self-organizing dynamics. Kelso and colleagues have extended and developed this type of phase-transition model to apply to a wide variety of cognitive domains, such as motor skill learning, speech perception, visual perception, and the dynamic coordination of activity among cortical areas of the brain (Bressler and Kelso 2001; Kelso 1995).

One of the key points relating to the dynamical approach is its emphasis on time. Traditional computational models are static in that they specify only a sequence of discrete states through which the system must pass. In contrast, dynamic-systems models specify how processes unfold in real time. As Tim van Gelder states, “Although all cognitive scientists understand cognition as something that happens *over* time, dynamicists see cognition as being *in* time, that is, as an essentially temporal phenomenon (van Gelder 1999a, p. 244). Van Gelder (1998) describes this contrast as one between *change* versus *state; geometry* versus *structure; structure in time* versus *static structure; time* versus *order; parallel* versus *serial;* and *ongoing* versus *input/output.*

Whereas computationalists focus primarily on discrete states and treat change as what happens when a system shifts from one discrete state to another, dynamicists focus on how a system changes state continuously in time. Dynamicists conceive of state changes geometrically, in terms of their position and trajectory in phase space, whereas computationalists focus on the internal formal or syntactic structure of combinatorial entities. Computationalists think of these structures as laid out statically (like snapshots), as either present all at once or not, and hence of cognition as the rule-governed transformation of one such static structure into another. For dynamicists, cognitive structures are laid out as temporally extended patterns of activity, and cognition is seen as the flow of complex temporal structures mutually and simultaneously influencing each other. Dynamicists are therefore interested in the timing (rates, periods, durations, synchronies) of processes, whereas computationalists have traditionally been interested only in the order of cognitive states. Moreover, computationalists tend to think of this order as being the serial or sequential progression of sense → perceive → think → act, whereas for dynamicists cognition unfolds as the continuous coevolution of acting, perceiving, imagining, feeling, and thinking. Finally, whereas computationalists think of cognitive processes as having an input-output structure—the system receives an input, proceeds through a sequence of internal operations, produces an output, and then halts—dynamicists think of processes as always ongoing, with no clear starting or end points. The goal is not to map an input at one time onto an output at a later time, but always to maintain appropriate change (van Gelder 1998).

Autonomous Systems

The dynamicist idea that cognitive processes are always ongoing with no clear starting or end points can be deepened by introducing the distinction between autonomous and heteronomous systems. *Autonomy* and *heteronomy* literally mean, respectively, self-governed and other-governed. A heteronomous system is one whose organization is defined by input-output information flow and external mechanisms of control. Traditional computational systems, cognitivist or connectionist, are heteronomous. For instance, a typical connectionist network has an input layer and an output layer; the inputs are initially assigned by the observer outside the system; and output performance is evaluated in relation to an externally imposed task. An autonomous system, however, is defined by its endogenous, self-organizing and self-controlling dynamics, does not have inputs and outputs in the usual sense, and determines the cognitive domain in which it operates (Varela 1979; Varela and Bourgine 1991).

In general, to specify any system one needs to describe its organization—the set of relations that defines it as the system it is. In complex systems theory, the term *autonomous* refers to a generic type of organization. The relations that define the autonomous organization hold between processes (such as metabolic reactions in a cell or neuronal firings in a cell assembly) rather than static entities. In an autonomous system, the constituent processes (i) recursively depend on each other for their generation and their realization as a network, (ii) constitute the system as a unity in whatever domain they exist, and (iii) determine a domain of possible interactions with the environment (Varela 1979, p. 55). The paradigm is a living cell. The constituent processes in this case are chemical; their recursive interdependence takes the form of a self-producing, metabolic network that also produces its own membrane; and this network constitutes the system as a unity in the biochemical domain and determines a domain of possible interactions with the environment. This kind of autonomy in the biochemical domain is known as *autopoiesis* (Maturana and Varela 1980). Figure 3.1 illustrates the basic organization required for autopoietic autonomy.

Autopoiesis is the paradigm case of biological autonomy for two reasons. It is empirically the best understood case, and it provides the core "biologic" of all life on Earth. To qualify as autonomous, however, a system does not have to be autopoietic in the strict sense (a self-producing bounded molecular system). An autopoietic system dynamically produces its own material boundary or membrane, but a system can be autonomous without having this sort of material boundary. The members of an insect colony, for example, form an autonomous social network, but the boundary is social and territorial, not material.

In exploring the notion of autonomy, we can take two complementary approaches—a top-down approach and a bottom-up one (Ruiz-Mirazo and Moreno 2004). Both approaches see autonomy as a relational, system-level property, but there is a critical difference between the two. Whereas the top-down approach focuses on the relational organization proper to autonomy, the bottom-up approach emphasizes the energetic and thermodynamic requirements for autonomy.

Varela takes the top-down approach in his 1979 book, *Principles of Biological Autonomy.*[3] In this work he defines an autonomous system as a system that has organizational closure (later called operational closure) (Varela 1979, pp. 55–60). Here closure does not mean that the system is

Figure 3.1. The basic autopoietic organization.

materially and energetically closed to the outside world (which of course is impossible). On the contrary, autonomous systems must be thermodynamically far-from-equilibrium systems, which incessantly exchange matter and energy with their surroundings. *Organizational closure* refers to the self-referential (circular and recursive) network of relations that defines the system as a unity, and *operational closure* to the reentrant and recurrent dynamics of such a system.[4] An autonomous system is always structurally coupled to its environment. Two or more systems are coupled when the conduct of each is a function of the conduct of the other. (In dynamic-systems language, the state variables of one system are parameters of the other system, and vice versa.) "Structural coupling" refers to the history of recurrent interactions between two or more systems that leads to a structural congruence between them (Maturana 1975; Maturana and Varela 1987, p. 75). Thus the state changes of an autonomous system result from its operational closure and structural coupling. The result of any state change is always further self-organized activity within the system, unless its closure is disrupted and it is no longer able to carry on its coupling, in which case it disintegrates. Systems described as autonomous in this sense abound throughout the living world—single cells, microbial communities, nervous systems, immune systems, multicellular organisms, ecosystems, and so on. Such systems need to be seen as sources of their own activity, specifying their own domains of interaction, not as transducers or functions for converting input instructions into output products. In other words, the autonomous character of these systems needs to be recognized.

The second, bottom-up approach to autonomy builds on these notions of organizational and operational closure, but tries to work out the energetic and thermodynamic requirements for the instantiation of "basic autonomy" in the physical world. From this perspective, basic autonomy is "the capacity of a system to manage the flow of matter and energy through it so that it can, at the same time, regulate, modify, and control: (i) internal self-constructive processes and (ii) processes of exchange with the environment" (Ruiz-Mirazo and Moreno 2004, p. 240). This capacity brings with it specific and demanding physical-implementation requirements: the system must have certain types of components, specifically a semipermeable active boundary (a membrane), an energy transduction/conversion apparatus (an energy currency such as adenosine triphosphate (ATP) in living cells, which transfers energy from chemical bonds to energy-absorbing reactions within the cell), and at least one type of component that controls and facilitates the self-construction processes (catalysts) (Ruiz-Mirazo and Moreno 2004, p. 252).

Figure 3.1 depicts the basic autopoietic organization for a living cell. A cell stands out of a molecular soup by creating the boundaries that set it apart from what it is not and that actively regulate its interactions with the environment. Metabolic processes within the cell construct these boundaries, but the metabolic processes themselves are made possible by those very boundaries. In this way, the cell emerges as a figure out of a chemical background. Should this process of self-production be interrupted, the cellular components no longer form a unity, gradually diffusing back into a molecular soup.

Figure 3.1 can be compared with Figure 3.2, which depicts the minimal form of organizational closure for a nervous system. Any nervous system operates according to a basic "neurologic," a pattern that continues and elaborates the biologic of autopoiesis. The fundamental

Figure 3.2. Organizational closure of the nervous system.

logic of the nervous system is to couple movement and a stream of sensory activity in a continuous circular fashion (Maturana and Varela 1987, pp. 142–176). Wherever movement is essential to a multicellular organism's mode of life, there is the corresponding development of a nervous system. A nervous system links sensory surfaces (sense organs and nerve endings) and effectors (muscles, glands) within the body. In this way it integrates the organism, holding it together as a mobile unity, as an autonomous sensorimotor agent.

This neurologic underlies all the variations on sensorimotor coordination found in the animal kingdom. In all animals, neuronal networks establish and maintain a sensorimotor cycle through which what the animal senses depends directly on how it moves, and how it moves depends directly on what it senses. No animal is a mere passive respondent; every animal meets the environment on its own sensorimotor terms. Merleau-Ponty recognized this crucial point in his first work, *The Structure of Behavior:*

> The organism cannot properly be compared to a keyboard on which the external stimuli would play and in which their proper form would be delineated for the simple reason that the organism contributes to the constitution of that form. When my hand follows each effort of a

> struggling animal while holding an instrument for capturing it, it is clear that each of my movements responds to an external stimulation; but it is also clear that these stimulations could not be received without the movements by which I expose my receptors to their influence . . . When the eye and the ear follow an animal in flight, it is impossible to say "which started first" in the exchange of stimuli and responses. Since all movements of the organism are always conditioned by external influences, one can, if one wishes, readily treat behavior as an effect of the milieu. But in the same way, since all the stimulations which the organism receives have in turn been possible only by its preceding movements which have culminated in exposing the receptor organ to the external influences, one could also say that the behavior is the first cause of the stimulations.
>
> Thus the form of the excitant is created by the organism itself, by its proper manner of offering itself to actions from outside. Doubtless, in order to be able to subsist, it must encounter a certain number of physical and chemical agents in its surroundings. But it is the organism itself—according to the proper nature of its receptors, the thresholds of its nerve centers and the movements of the organs—which chooses the stimuli in the physical world to which it will be sensitive. "The environment *(Umwelt)* emerges from the world through the actualization or the being of the organism—[granted that] an organism can exist only if it succeeds in finding in the world an adequate environment." This would be a key board which moves itself in such a way as to offer—and according to variable rhythms—such or such of its keys to the in itself monotonous action of an external hammer. (1963, p. 13)[5]

This passage clearly expresses an autonomy perspective. Organisms display patterns of behavior that require us to see them as autonomous. Varela tries to characterize this autonomy at an abstract level in terms of a generic dynamic pattern or form, namely, organizational and operational closure. Hence Varela gives us his "Closure Thesis," which states, "Every autonomous system is organizationally closed" (Varela 1979, p. 58).[6]

Figure 3.1 illustrates the minimal form this closure takes for life at the single-cell level, and Figure 3.2 illustrates the minimal form it takes for the nervous system. Whereas autopoietic closure brings forth a minimal "bodily self" at the level of cellular metabolism, sensorimotor closure produces a "sensorimotor self" at the level of perception and action. In the one case the passage from network closure to selfhood (and correlative otherness) happens at the level of an active semipermeable boundary or membrane, which regulates interaction with the outside environment. In the other case it happens at the level of behavior and intentional action. In both cases we see the co-emergence of inside and outside, of selfhood and a correlative world or environment of otherness, through the generic mechanism of network closure (autonomy) and its physical embodiment (Varela 1997a; see also Moreno and Barandiaran 2004).[7]

In addition to these cellular and sensorimotor forms of selfhood, other forms of selfhood arise from other organizationally and operationally closed systems. The immune system, for instance—understood as an autonomous immune network that establishes a coherent somatic identity for the organism, rather than as a mere mechanism of defense—brings forth a dynamic, somatic identity at a distributed cellular and molecular level (Coutinho 2003; Varela and Coutinho 1991). The animate form of our living body is thus the place of intersection for numerous emergent patterns of selfhood and coupling. Whether cellular, somatic, sensorimotor, or neurocognitive, these patterns derive not from any homuncular self or agent inside the system organizing it or directing it, but from distributed networks with operational closure. In Varela's image, our organism is a meshwork of "selfless selves," and we are and live this meshwork (Varela 1991; Varela and Cohen 1989).

Let me forestall an objection that might arise at this point. The nervous system is clearly embedded in the body of the organism, and the organism in its environment (Chiel and Beer 1997). This fact seemingly contradicts the statement that the nervous system is an autonomous system and that the organism is an autonomous agent. The thought here would be that the operation of the nervous system loops through the body (via sensory and motor surfaces), and therefore it is not possible that the nervous system has operational closure (that the product of every process within the system stays within that system). Similarly, because the bodily activity of the organism loops through the environment (motor activity affects the sensory stimulation one receives back from the environment), the organism cannot have an operationally closed dynamics.

A number of points, both methodological and epistemological, need to be made in reply. The first point is that, strictly speaking, system, autonomy, and heteronomy are heuristic notions—they are cognitive aids or guides in the scientific investigation and characterization of observable phenomena and patterns of behavior. As heuristic notions, they (implicitly) refer back to and implicate the interpretive and explanatory stance of an observer (or observer community). What counts as the system in any given case, and hence whether it is autonomous or heteronomous, is context-dependent and interest-relative. For any system it is always possible to adopt a heteronomy or external-control perspective, and this can be useful for many purposes. Nevertheless, this stance does not illuminate—and indeed can obscure—certain observable patterns of behavior, namely, patterns arising from the system's internal dynamics rather than external parameters. An organism dynamically produces and maintains its own organization as an invariant through change, and thereby also brings forth its own domain of interaction. (Although living organisms are paradigmatic in this regard, nothing apparently rules out the possibility of artificial autonomy.) A heteronomy perspective does not provide an adequate framework to investigate and understand this phenomenon; an autonomy perspective is needed.

The second point is that in any given case or for any candidate system we need to distinguish between, on the one hand, the operation of the system as such, which is a function of both its organization (the set of relations that defines it as a system) and physical structure, and, on the other hand, its performance in relation to whatever wider context in which it is observed. For example, if we wish to characterize the organization and operation of the nervous system as a finite neuronal network, then we need to characterize the nervous system as organizationally and operationally closed, such that any change of activity in a neuron (or neural assembly) always leads to a change of activity in other neurons (either directly through synaptic action or indirectly through intervening physical and chemical elements). Sensory and effector neurons are no exception because any change in the one leads to changes in the other, such that the network always closes back upon itself, regardless of intervening elements (Maturana and Varela 1980, p. 127). Nevertheless, the domain of states available to the nervous system (as an operationally closed network) is clearly a function of its history of interactions with the rest of the body (and the environment). Hence, besides characterizing the nervous system's operation as a closed network, we need to characterize its performance in its structural coupling with the rest of the body (and the environment). Similarly, to characterize the organism as a finite cellular or multicellular entity, we need to characterize it as an organizationally and operationally closed system. At the same time, we need to characterize the organism's performance or behavior in its structural coupling with the environment.

We can also shift perspectives and characterize the nervous system as a heteronomous system—that is, as a component system with various functions defined in relation to the organism (such as registering sensory changes or guiding movement). Notice, however—and this is the third point—that in so shifting perspectives we are *ipso facto* no longer talking about the same system. The system with which we are now concerned is no longer the nervous system as a finite neuronal network, but rather the larger system of the organism (in which the nervous system is seen as a component). Similarly, we can also shift perspectives and characterize the organism as a heteronomous system subject to the control of the environment (for instance, other organisms). Once again, in thus shifting perspectives we are *ipso facto* no longer dealing with the same system. The system now is the larger system of organism-plus-environment, not the organism as a finite cellular or multicellular entity.[8]

These considerations show us that there is no inconsistency between characterizing the nervous system and organism as autonomous and emphasizing their somatic and environmental embeddedness. We do, however, have to keep our logical and conceptual accounts clear, so that we know which explanatory heuristic is in play at any given time. In any case, for the enactive approach it is the autonomy perspective on natural cognitive agents that remains the reference point for understanding mind and life, not a predefined input-output task strucure.

Information and Meaning

Adopting an autonomy perspective also brings with it a certain way of thinking about semantic information or meaning. For enactive theorists, information is context-dependent and agent-relative; it belongs to the coupling of a system and its environment. What counts as information is determined by the history, structure, and needs of the system acting in its environment.

According to the received view in cognitive science, in order to explain cognitive abilities we need to appeal to information-bearing states inside the system. Such states, by virtue of the semantic information they carry about the world, qualify as representations. Cognitivists conceive of these representations as symbols in a computational "language of thought," and connectionists as constrained patterns of network activity corresponding to phase space "attractors" (regions of phase space toward which all nearby trajectories converge). In either case there is a strong tendency to adopt an objectivist conception of representation: representations are internal structures that encode context-independent information about the world, and cognition is the processing of such information.

This objectivist notion of information presupposes a heteronomy perspective in which an observer or designer stands outside the system and states what is to count as information (and hence what is to count as error or success in representation). Information looks different from an autonomy perspective. Here the system, on the basis of its operationally closed dynamics and mode of structural coupling with the environment, helps determine what information is or can be.

A neurobiological example can help to illustrate these ideas.[9] Certain kinds of cortical neurons are often described as feature detectors because they respond preferentially (fire above their base rate) to various types of stimuli, such as edges, lines, and moving spots. Such neurons are identified by recording their individual activity with a microelectrode and determining the sensory stimulus to which the neuron is most sensitive. Such neurons are said to "represent" features of objects and to make that information available for futher processing by various systems in the brain. This view lies behind the standard formulation of the so-called binding problem. This problem concerns how distinct features (shape, color, motion), as represented by cell populations in spatially distributed and functionally segregated neural pathways, can be bound together to form a complete and accurate representation of the object (so that the right shapes go with the right colors and motions). This way of thinking about the brain treats it as a heteronomous system: object features outside the organism provide informational inputs to the brain, and the brain's information processing task is to arrive at an accurate representation of the objective world and produce an adaptive motor output.

From an autonomy perspective, it is crucial to distinguish between information about stimuli as they are defined by an observer and information in the sense of what meanings the stimuli have for the animal. Only the latter play a significant role in the brain's operation. The notion of an object "feature" is defined by an observer who stands outside the system, has independent access to the environment, and establishes correlations between environmental features and neuronal responses. The animal's brain has no access to features in this sense (and *a fortiori* has no access to any mapping from features to neuronal responses). As Freeman explains, "In the view from neurodynamics, neurons that respond to edges, lines, and moving spots are manifesting the local topological properties of neuronal maps, which extract local time and space derivatives in automatic preprocessing for spatial and temporal contrast enhancement. No objects or features are manifested at the level of the single neuron, assuredly not those used by an observer" (Freeman 1995, p. 54). From an autonomy perspective, individual neurons do not detect objectively defined features. Rather, assemblies of neurons make sense of stimulation by constructing meaning, and this meaning arises as a function of how the brain's endogenous and nonlinear activity compensates for sensory perturbations. From this perspective, the feature-binding problem is not the brain's problem, but the brain theorist's problem; it is an artifact of a certain way of looking at the brain. Freeman's description of the alternative view, based on looking at the brain as an autonomous system operating according to nonlinear causality, is well worth quoting here:

> In this view the experimenter trains a subject to co-operate through the use of positive and negative reinforcement, thereby inducing a state of expectancy and search for a stimulus, as it is conceived by the subject. When the expected stimulus arrives, the activated receptors transmit pulses to the sensory cortex, where they elicit the construction by nonlinear dynamics of a macroscopic, spatially coherent oscillatory pattern that covers an entire area of sensory cortex . . . It is observed by means of the electroencephalogram (EEG) from electrode arrays on all the sensory cortices . . . It is not seen in recordings from single neuron action potentials, because the fraction of the variance in the single neu-

> ronal pulse train that is covariant with the neural mass is far too small, on the order of 0.1 percent.
>
> The emergent pattern is not a representation of a stimulus . . . It is a state transition that is induced by a stimulus, followed by a construction of a pattern that is shaped by the synaptic modification among cortical neurons from prior learning. It is also dependent on the brain stem nuclei that bathe the forebrain in neuromodulatory chemicals. It is a dynamic action pattern that creates and carries the meaning of the stimulus for the subject. It reflects the individual history, present context, and expectancy, corresponding to the unity and wholeness of intentionality. Owing to dependence on history, the patterns created in each cortex are unique to each subject. (Freeman 1999b, pp. 149–150)

The distinction between autonomous meaning-construction and heteronomous information processing needs to be placed in the broader context of the embodied dynamicist way of thinking about information. To explain this way of thinking, it will be helpful to go back to ideas introduced by Howard Pattee (1977). Pattee made an important distinction between two modes of description of a complex system—the *linguistic mode*, which describes the system in terms of discrete, rate-independent, symbolic elements, and the *dynamical mode*, which describes the system in terms of continuous, rate-dependent processes, and thus explicitly includes the flow of time. Pattee raised the following question: "How do we know we are not intepreting certain structures as descriptions, only because we recognize them as consistent with rules of one of our own languages?" (1977, p. 262). In other words, how do we know our linguistic descriptions are not simply observer-relative, but rather correspond to symbolic structures that belong to the system itself and play a role in its operation? And he answered: "we must further restrict our model of a complex system to remove the case of the external observer reading a message that is not really in the system itself. This restriction is achieved by requiring that a complex system must read and write its own messages" (1977, p. 262).

Pattee's example is a living cell. When we describe DNA triplets as "coding for" amino acids, we employ the linguistic mode of description. Which amino acid a given DNA triplet specifies is supposed to be rate-independent—it does not matter how fast the triplet is "read" in the course of protein synthesis. It is also supposed to be arbitrary, in the sense that "[i]t is hard to see why a code in which GGC means glycine and AAG means lycine is either better or worse than one in which the meanings are reservsed" (Maynard Smith 1986, p. 19). According to Pattee, the linguistic mode of description in this case is not observer-relative because the cell is a self-describing system that "reads and writes its own messages." The writing of its own messages corresponds to DNA replication (the production of a complement of the original DNA molecule through a template); the reading of its own messages corresponds to protein synthesis (DNA "transcription" to RNA and RNA "translation" to protein).

Pattee then makes a number of crucial points. First, for the code to be read there must ultimately be a transduction or conversion within the cell from the linguistic mode to the dynamical mode. This conversion occurs when the rate-independent linear array of amino acids folds to become a three-dimensional enzyme. Within the life cycle of the cell, there is thus a transformation from the enzyme as something designated in the genome to the enzyme as an operational component of metabolism. Second, this transformation (the protein folding) is not itself linguistically described in the cell, but rather happens according to physical law (under the higher-order constraint of the DNA-specificed amino acid sequence). Third, if the transformation were linguistically described, the speed and precision with which it is accomplished would be considerably compromised. Pattee's conclusion is that "we would not expect a complete formal description or simulation of a complex system to adapt or function as rapidly or reliably as the partially self-describing, tacit dynamic system it simulates" (1977, p. 264).

Pattee emphasizes the complementarity of the linguistic and dynamical modes of description, but also suggests that symbolic information emerges from and acts as a constraint on dynamics. This idea is important for embodied dynamicism and the enactive approach. Let us return to the example of the cell. In general, nucleotide triplets are capable of predictably specifying an amino acid if and only if they are properly embedded in the cell's metabolism, that is, in a multitude of enyzmatic regulations in a complex biochemical network. This network has a chicken-and-egg character at several levels. First, proteins can arise only from a DNA/RNA "reading" process, but this process cannot happen without proteins. Second, the DNA "writing" and "reading" processes must be properly situated within the intracellular environment, but this environment is a result of those very processes. Finally, the entire cell is an autopoietic system—that is, an autonomous system defined by an operationally closed network of molecular processes that simultaneously both produces and realizes the cell concretely in the physical space of its biochemical components.

Now, when we employ the linguistic mode of description and state that DNA/RNA "codes" for proteins, we restrict our focus to one particular sequence of this overall circular causality. We abstract away from the many intervening and necessary causal steps in the actual dynamic process of protein synthesis, and we bracket out the essential participation of many other molecular elements (such as RNA polymerase enzymes, and positive and negative regulatory proteins). We "thus reduce our description to a skeleton that associates a certain part of a nucleic acid with a certain protein segment. Next we observe that this kind of simplified description of an actual dynamic process is a useful one in following the sequences of reproductive steps from one generation to the other, to the extent that the dynamic process stays stable (i.e., the kinds of dynamics responsible for bonding, folding, and so on) . . . A symbolic explanation, such as the description of some cellular components as genes, betrays the emergence of certain *coherent patterns of behavior to which we choose to pay attention*" (Varela 1979, p. 75). It is the emergence of such coherent dynamic patterns that underwrites the symbolic informational level of description: "An object or event is a symbol only if it is a token for an abbreviated nomic chain that occurs *within the bounds of the system's organizational closure.* In other words, whenever the system's closure determines certain regularities in the face of internal or external interactions and perturbations, such regularities can be abbreviated as a symbol, usually the initial or terminal element in the nomic chain" (Varela 1979, pp. 79–80). Thus, when we talk about DNA "coding" for proteins we are not referring to a special type of symbolic causal relation or a special type of intrinsically informational molecule that rises above the dynamic fray. Rather, we are abbreviating a lengthy but remarkably stable dynamic pattern of biochemical events. It is precisely the stability and predictability of the entire pattern that allows us to telescope it in a linguistic mode of description by treating nucleotide triplets as in effect "standing for" amino acids.

This mode of description is unobjectionable (and has heuristic value) as long as it is remembered that the genetic code is no more than a rule of causal specificity based on the fact that cells use nucleic acids as templates for the primary structure (amino acid sequence) of proteins (Godfrey-Smith 2000b; Thompson 1997). Yet it is unacceptable to say that DNA contains the information for phenotypic design, because this statement attributes an intrinsic semantic-informational status to one particular type of component and thereby divests this component of its necessary embedding in the dynamics of the autopoietic network. It is this network in its entirety that specifies the phenotypic characteristics of a cell, not one of its components, and it is this network as a whole that serves as the precondition and causal basis of DNA replication ("writing") and protein synthesis ("reading") (see Moss 1992). Information is not intrinsic to the static linear array of the DNA sequence, but is rather dynamically constituted in and by the cell as an autopoietically organized, three-dimensional entity—by the cell as a *body*. In summary, the linguistic mode is emergent from the dynamical mode, and information exists only as dynamically embodied.

With these points having been made, we return to the difference between autonomous meaning-construction and heteronomous information processing. Information is formed within a context rather than imposed from without. Gregory Bateson used to say, "information is a difference that makes a difference" (Bateson 1972, p. 315). We could elaborate this insight by saying that information, dynamically conceived, is the making of a difference that makes a difference for some*body* somewhere (see Oyama 2000b). Information here is understood in the sense of *informare,* to form within (Varela 1979, p. 266). An autonomous system becomes informed by virtue of the meaning formation in which it participates, and this meaning formation depends on the way its endogenous dynamics specifies things that make a difference to it (Kelso and Kay 1987; Turvey and Shaw 1999).

For another example we can return to the finger coordination study of Haken, Kelso, and Bunz (1985). There the switching from in-phase to anti-phase happens without any command from a motor program; rather, it occurs spontaneously as an emergent property of the system's self-organizing dynamics. The collective variable or order parameter of relative phase is informational in the sense that it specifies coherent patterns or relations that inform the system and that can be physically or physiologically realized in multiple ways. As Kelso explains:

> Instead of treating dynamics as ordinary physics using standard biophysical quantities such as mass, length, momentum, and energy, our *coordination or pattern dynamics* is informational from the very start. The order parameter, ϕ [relative phase], captures the *coherent relations* among different kinds of things. Unlike ordinary physics, the pattern dynamics is context dependent: the dynamics are valid for a given biological function or task, but largely independent of how this function is physiologically implemented. Thus, if we accept that the same order parameter, ϕ, captures coherent spatiotemporal relations among different kinds of things, and the same equations of motion describe how different coordination patterns form, coexist, and change, it seems justified to conclude that order parameters in biological systems are functionally specific, context-sensitive *informational* variables; and that the coordination dynamics are more general than the particular structures that instantiate them.
>
> Notice, coordination dynamics is not trapped (like ordinary physics) by its (purely formal) syntax. Order parameters are semantic, relational quantities that are intrinsically meaningful to system functioning. What could be more meaningful to an organism than information that specifies the coordinative relations among its parts or between itself and the environment? This view turns the mind-matter, information-dynamics interaction on its head. Instead of treating dynamics as ordinary physics and information as a symbolic code acting in the way that a program relates to a computer, dynamics is cast in terms that are semantically meaningful. (Kelso 1995, p. 145)

Let me connect these points to the autonomy perspective. As we have seen, from the autonomy perspective a natural cognitive agent—an organism, animal, or person—does not process information in a context-independent sense. Rather, it brings forth or enacts meaning in structural coupling with its environment. The meanings of an autonomous system's states are formed within *(informare)* the context of the system's dynamics and structural coupling. Therefore, if we wish to continue using the term *representation*, then we need to be aware of what sense this term can have for the enactive approach. Representational "vehicles" (the structures or processes that embody meaning) are temporally extended patterns of activity that can crisscross the brain-body-world boundaries, and the meanings or contents they embody are brought forth or enacted in the context of the system's structural coupling with its environment.

Another way to make this point would be to say that autonomous systems do not operate on the basis of internal representations in the subjectivist/objectivist sense. Instead of internally representing an external world in some Cartesian sense, they enact an environment inseparable from their own structure and actions (Varela, Thompson, and Rosch 1991, p. 140). In phenomenological language, they constitute (disclose) a world that bears the stamp of their own structure. As Merleau-Ponty puts it, quoting Goldstein, in the passage cited earlier: "the environment emerges from the world through the being or actualization of the organism." In the case of animal life, the environment emerges as a sensorimotor world through the actualization of the organism as a sensorimotor being. The organism is a sensorimotor being thanks to its nervous system. The nervous system connects anatomically distant sensory and motor processes, subsuming them in operationally closed sensorimotor networks. Through their coherent, large-scale patterns of activity these networks establish a sensorimotor identity for the animal—a sensorimotor self. In the same stroke, they specify what counts as "other," namely, the animal's sensorimotor world.

This idea of a sensorimotor world—a body-oriented world of perception and action—is none other than von Uexküll's original notion of an *Umwelt*. An *Umwelt* is an animal's environment in the sense of its lived, phenomenal world, the world as it presents itself to that animal thanks to its sensorimotor repertoire: "all that a subject perceives becomes his perceptual world and all that he does, his effector-world. Perceptual and effector worlds together form a closed unit, the Umwelt" (von Uexküll 1957, p. 6). The logic of this co-emergence is depicted in Figure 3.3.

In this figure, information is the intentional relation of the system to its milieu, established on the basis of the system's autonomy (organizational-operational closure). One of the main scientific tasks for embodied dynamicism and the enactive approach is to explain how the

Figure 3.3. Co-emergence of autonomous selfhood and world.

pattern dynamics of brain, body, and behavior are informational in this sense (see Kelso 1995, p. 288).

Emergent Processes

Another key idea of the enactive approach that needs elaboration here is the idea of *emergence* or *emergent processes*. In complex systems theory, an emergent process is one that results from collective self-organization. An emergent process belongs to an ensemble or network of elements, arises spontaneously or self-organizes from the locally defined and globally constrained or controlled interactions of those elements, and does not belong to any single element. The enactive approach builds on this notion of emergence but reformulates it as "dynamic co-emergence," in which part and whole co-emerge and mutually specify each other.

Let me first introduce emergence in the complex systems sense. A standard example of this kind of emergence is the formation of "Bénard cells," a dynamic geometrical pattern of fluid flow, in liquids or gases subject to an imposed temperature gradient (see Kelso 1995, pp. 7–8; Solé and Goodwin 2000, pp. 13–17). The emergence of Bénard cells can be seen in the behavior of cooking oil in a frying pan. Applying heat to the pan increases the temperature difference between the cooler layer of oil at the top and the hotter layer of oil at the bottom. When the temperature difference between top and bottom is small, there is no large-scale or global motion of the oil, but eventually when the difference becomes large enough instability occurs and the liquid starts to roll in an orderly fashion known as convection rolls. In other words, the system undergoes a state transition, described mathematically as a bifurcation, as the new self-organizing behavior and spatial structures of convection rolls emerge. As the temperature gradient is increased still further, the convection rolls undergo another transition or bifurcation and give rise to an array of hexagonal up-and-down flow patterns called Bénard cells.

This example illustrates several basic points about collective self-organization and dynamic emergence. The temperature gradient is the *control parameter* for the transitions or bifurcations. It leads the system through a variety of possible states but does not dictate, prescribe, or code for the emerging flow patterns. Nor is there any homunculus or program inside the system determining those patterns. In Kelso's words: "Such spontaneous pattern formation is exactly what we mean by self-organization: the system organizes itself, but there is no 'self,' no agent inside the system doing the organizing" (Kelso 1995, p. 8). The *order parameter* of the system is the amplitude of the convection rolls. It is created by the interactions of the fluid molecules, but at the same time it governs or constrains their behavior by drastically reducing the immense number of degrees of freedom of motion that the individual molecules would otherwise have.

Emergence through collective self-organization thus has two aspects. One is local-to-global determination, as a result of which novel macrolevel structures and processes emerge. The other is global-to-local determination whereby global structures and processes constrain local interactions. These global-to-local influences do not take the same form as the local-to-global ones: they typically manifest themselves through changes in control parameters (the temperature gradient in the example above) and boundary conditions rather than through changes to the individual elements (the fluid molecules). Coherent and ordered global behaviors, which are described by collective variables or order parameters, constrain or govern the behavior of the individual components, entraining them so that they no longer have the same behavioral alternatives open to them as they would if they were not interdependently woven into the coherent and ordered global pattern. At the same time, the behavior of the components generates and sustains the global order. This two-sided or double determination is known as circular causality (Haken 1983).

Emergence and circular causality are crucially important in the context of neurodynamics. Neuroscience indicates that cognition, emotion, and action require the transient integration of numerous, widely distributed, and constantly interacting brain regions and areas. An outstanding question for neuroscience today is to determine the mechanisms of this large-scale integration. From a dynamic-systems perspective, large-scale integration corresponds to the formation of transient dynamic links between widespread neural populations (Varela et al. 2001). On the one hand, large-scale dynamic patterns emerge from distributed local neuronal activities. On the other hand, large-scale patterns constrain these local activities. According to a number of theorists, dynamic instability or metastability is crucial to large-scale integration because it permits a flexible repertoire of global states without the system becoming trapped in any one particular state.[10]

Emergence and circular causality can also be illustrated by neurodynamical studies of epilepsy (Thompson and Varela 2001). Focal epileptic seizures originate in specific parts of the cortex; they can remain confined to those areas or spread to other parts of the brain. Their clinical manifestations depend on the cortical area in which they originate, how widely they spread, and how long they last. Local epileptic activity can modify the subject's mental competencies and give rise to various kinds of mental experiences, such as visual or auditory illusions and hallucinations, and memory phenomena involving the vivid actual recall of a past event or déjà-vu illusions. These mental phenomena can also be induced by direct electrical stimulation of the temporal lobe in epileptic patients, as classically described by Wilder Penfield (1938). Thus local neuronal activity at the level of an epileptogenic zone can produce large-scale effects, eventually influencing the global level of a moment of consciousness. This is a case of local-to-global emergence in the brain.

The converse—global-to-local influence—though less documented and more controversial, seems also to be the case. The basic idea is that cognitive activity, which reflects large-scale integration in the brain, can affect local activity. For example, the subject can voluntarily affect local epileptic activity, as indicated by numerous patient reports and a few clinically reported cases (see Le Van Quyen and Petitmengin 2002). As long ago as 1954, Penfield and Jasper described the blocking of a parietal seizure by the initiation of a complex mathematical calculation (Penfield and Jasper 1954), and recently more extensive observations have confirmed such cognitive influences (Schmid-Schonbein 1998). We can assume that such intervention is possible because the epileptogenic zones are embedded in a complex network of other brain regions that actively participate in the large-scale integration underlying cognitive acts. It also seems reasonable to assume that these global patterns of integration can influence local events, including eventually the local epileptogenic zones, whose activity can thus be taken as an indication of the global influence.

Experimental work by Michel Le Van Quyen, Francisco Varela, and their colleagues provides evidence for such global-to-local influence in the case of a patient with an unusually focal and stable occipitotemporal epileptic discharge.[11] This patient showed no evidence of cognitive impairment and was willing to participate in simple cognitive tasks of visual and auditory discrimination. For the visual task, he was asked to press a button when the target stimulus appeared, but not when the two other nontarget stimuli were shown. The temporal intervals between successive discharges of the epileptic spike pattern were analyzed. Dynamical inspection (in the form of a phase space known as a first-return map) showed that the distribution of the intervals followed a particular kind of unstable dynamic pattern. The spikes displayed a distinct periodic activity for a short time before they diverged away along another unstable direction, a kind of dynamic pattern known as an unstable periodic orbit. Furthermore, this activity covaried with the specific mental state of the patient during the perceptual task and appeared to be modulated by the gamma frequency (30–70 Hertz) activity associated with his cognitive states. (Gamma frequency activity is widely reported to be associated with a variety of cognitive processes, including attention, perception, and memory.) These findings suggest that the patient's act of perception contributed in a highly specific way to "pulling" the epileptic activities toward particular unstable periodic orbits. Such global-to-local influence mobilized by cognitive activity might open up possibilities for cognitive strategies of control of epileptic seizures (Le Van Quyen and Petitmengin 2002).

Let me conclude this chapter by linking these ideas about circular causality and emergence back to autonomy. An autonomy perspective brings with it a certain way of thinking about emergence. What emerges in the case of an autonomous system such as a cell is a self-producing entity that also brings forth its own domain of interactions (see Figure 3.3). This sort of emergence takes a major step beyond dynamic pattern formation in physical dissipative systems:

> [A]lthough the phenomenon of self-organization always involves the generation and maintenance of a global (or high-level) pattern or correlation that constrains the (low-level) dynamics of the components of the system, in standard dissipative structures this occurs only provided that the system is put under the appropriate boundary conditions. If those (externally controlled) conditions are changed (in particular, if the input of matter or energy is outside a certain range), the self-organizing dynamic vanishes. Therefore, there is an important difference between the typical examples of "spontaneous" dissipative structures and real autonomous systems: in the former case, the flow of energy and/or matter that keeps the system away from equilibrium is not controlled by the organization of the system (the key boundary conditions are externally established, either by the scientist in the lab or by some natural phenomenon that is not causally dependent on the self-organizing one), whereas in the latter case, the constraints that actually guide energy/matter flows from the environment through the constitutive processes of the system are endogenously created and maintained. (Ruiz-Mirazo and Moreno 2004, p. 238)

An autonomous system, such as a cell or multicellular organism, is not merely self-maintaining, like a candle flame; it is also self-producing and thus produces its own self-maintaining processes, including an active topological boundary that demarcates inside from outside and actively regulates interaction with the environment. In the single-cell, autopoietic form of autonomy, a membrane-bounded, metabolic network produces the metabolites that constitute both the network itself and the membrane that permits the network's bounded dynamics. Other autonomous systems have different sorts of self-constructing processes and network topologies. Whether the system is a cell, immune network, nervous system, insect colony, or animal society, what emerges is a unity with its own self-producing identity and domain of interactions or milieu, be it cellular (autopoiesis), somatic (immune networks), sensorimotor and neurocognitive (the nervous system), or social (animal societies).

Dynamic co-emergence best describes the sort of emergence we see in autonomy. In an autonomous system, the whole not only arises from the (organizational closure of) the parts, but the parts also arise from the whole.[12] The whole is constituted by the relations of the parts, and the parts are constituted by the relations they bear to one another in the whole. Hence, the parts do not exist in advance, prior to the whole, as independent entities that retain their identity in the whole. Rather, part and whole co-emerge and mutually specify each other.

Biological life, seen from the perspective of autopoiesis, provides a paradigm case of dynamic co-emergence. A minimal autopoietic whole emerges from the dynamic interdependence of a membrane boundary and an internal chemical reaction network. The membrane and reaction network (as well as the molecules that compose them) do not pre-exist as independent entities. Rather, they co-emerge through their integrative, metabolic relation to each other. They produce and constitute the whole, while the whole produces them and subordinates them to it. We will come back to this paradigm case of dynamic co-emergence in Part II.

Index