THE CONTINUITY OF MIND

Michael Spivey

OXFORD PSYCHOLOGY SERIES

The Continuity of Mind

RECENT TITLES IN THE OXFORD PSYCHOLOGY SERIES

Editors

Mark D'Esposito	Daniel Schacter
Jon Driver	Anne Treisman
Trevor Robbins	Lawrence Weiskrantz

22. Classification and cognition
    W. K. Estes
23. Vowel perception and production
    B. S. Rosner and J. B. Pickering
24. Visual stress
    Arnold Wilkins
25. Electrophysiology of mind: event-related brain potentials and cognition
    Edited by Michael D. Rugg and Michael G. H. Coles
26. Attention and memory: an integrated framework
    Nelson Cowan
27. The visual brain in action
    A. David Milner and Melvyn A. Goodale
28. Perceptual consequences of cochlear damage
    Brian C. J. Moore
29. Binocular vision and stereopsis
    Ian P. Howard and Brian J. Rogers
30. The measurement of sensation
    Donald Laming
31. Conditioned taste aversion: memory of a special kind
    Jan Bures, Federico Bermúdez-Rattoni, and Takashi Yamamoto
32. The developing visual brain
    Janette Atkinson
33. The neuropsychology of anxiety: an enquiry into the functions of the septo-hippocampal system (second edition)
    Jeffrey A. Gray and Neil McNaughton
34. Looking down on human intelligence: from psychometrics to the brain
    Ian J. Deary
35. From conditioning to conscious recollection: memory systems of the brain
    Howard Eichenbaum and Neal J. Cohen
36. Understanding figurative language: from metaphors to idioms
    Sam Glucksberg
37. Active vision: the psychology of looking and seeing
    John M Findlay and Iain D Gilchrist
38. The science of false memory
    C. J. Brainerd and V. F. Reyna
39. The case for mental imagery
    Stephen M. Kosslyn, William L. Thompson, and Giorgio Ganis
40. The continuity of mind
    Michael Spivey

The Continuity of Mind

Michael Spivey

OXFORD
UNIVERSITY PRESS

2007

OXFORD
UNIVERSITY PRESS

Oxford University Press, Inc., publishes works that further
Oxford University's objective of excellence
in research, scholarship, and education.

Oxford New York
Auckland Cape Town Dar es Salaam Hong Kong Karachi
Kuala Lumpur Madrid Melbourne Mexico City Nairobi
New Delhi Shanghai Taipei Toronto

With offices in
Argentina Austria Brazil Chile Czech Republic France Greece
Guatemala Hungary Italy Japan Poland Portugal Singapore
South Korea Switzerland Thailand Turkey Ukraine Vietnam

Published by Oxford University Press, Inc.
198 Madison Avenue, New York, New York 10016

www.oup.com

Library of Congress Cataloging-in-Publication Data
Spivey, J. M.
The continuity of mind / by Michael Spivey.
p. cm. — (Oxford psychology series; no. 40)
ISBN-13 978-0-19-517078-8
ISBN 0-19-517078-4
1. Cognition. I. Title. II. Series.
BF311.S6695 2006
153—dc22 2006005937

9 8 7 6 5 4 3 2 1

Printed in the United States of America
on acid-free paper

For Gramma Donna

This page intentionally left blank

Foreword

This book marks a major step forward in cognitive science, an effective way of thinking about minds and brains that isn't just another computer metaphor. Many of us have been looking for such a step, but where would it come from? One promising possibility was dynamical systems theory, which indeed is basic to Michael Spivey's argument here. Until now, however, dynamical systems have had little to say about genuinely cognitive achievements such as language, categorization, or thought. Neural nets have been another promising possibility (one that also plays a role here), but most of them are still essentially step-by-step computer models indifferent to the properties of real neurons that live in real time. On the empirical side there have been many ingenious new methods and exciting new findings in recent years, but until now no coherent theory has emerged to hold them all together. How could any theory deal with so much complexity?

Here's how. First, any such theory will have to establish its own units of analysis. What could those units be? They can't just be *responses*: The early behaviorists took responses as far as they would go, which wasn't very far. It also won't do to start with *information*, the vehicle that made cognitive psychology possible a generation ago. Of course, it's still true that brains process information, but saying so is no longer revolutionary or even very helpful. Nor can the basic units be *single neurons*: that soon leads to "grandmother cells," implausible for many reasons. Spivey's proposal here—a seriously expanded version of dynamical systems theory with many original twists—is based instead on trajectories through the state space of the human brain. His insistence that those trajectories must be continuous

has led him to new insights over a surprisingly broad range of cognitive phenomena.

But what is a state space? What sorts of things move through state spaces? What does it mean to assert that those movements are continuous? Taking the last question first, "continuity" means that movements away from a given brain state are always to an adjacent state and always take real time, a time during which much can happen. Speech perception provides a convenient example. Although a spoken word is not fully defined until its last syllable ends, the process of understanding it starts much earlier. *Candle* and *candy*, for example, both begin with *can*. Spivey's ingenious eye movement studies show that a listener presented with one of these words will actively consider both those possibilities at first, making a commitment only later as more information arrives. The moral here is that word representations—indeed, all mental representations—are probabilistic and overlapping rather than sharply bounded. The brain is "hungry" for information, always using whatever it has and looking for more.

These characteristics have implications for the theory's units of analysis. A representation capable of overlapping widely and probabilistically with other representations must involve a large number of neurons, some of which are active at a given moment while others are not. Such collections of neurons are *distributed representations* or *population codes*. Their interwoven patterns of activation are what produce the effects we observe.

Important as they are, population codes are not the ultimate units of analysis. To provide a richer description of the brain's activity, Spivey uses a *multidimensional state space*. Each brain neuron corresponds to one dimension of that space, which thus has a billion or so dimensions. At any given moment, the total state of brain activity corresponds to a single point in the space. Changes in that activity over time then produce *trajectories* through the space. Regions of the space to which many trajectories go (and where they sort of stay) are called *attractor basins*. In many contexts a given attractor basin corresponds to a fully developed percept—to a word understood, a face recognized, a stable perceived version of the Necker cube. The attractors are thus very important, but Spivey is even more interested in the trajectories themselves. The basic units of his thinking are events, not states.

The Continuity of Mind is not an easy book, but its organization is clear. After the introduction (chapter 1), Spivey devotes three chapters to intellectual tools that the rest of the argument will require. The first of these, chapter 2, reviews the logic of state space representations. Chapter 3 surveys such diverse but relevant paradigms as reaction time, MEG, ERP, EEG, single-cell recording, repetitive rhythmic motor tasks, 3D motion capture, and especially eye movements. Eye tracking is Spivey's favorite paradigm, not only because he has worked on it so effectively himself but also because it is surprisingly good at revealing rapid mental activity that occurs outside of consciousness. Then comes the third conceptual-tool chapter, chapter 4, which is specifically designed "to gently walk the reader through some of the mathematics of a few simple demonstrations of dynamical systems." It does help.

With these conceptual tools in hand, Spivey sets out to show how his continuity assumption addresses the major issues of contemporary cognitive science. The first of those issues is modularity à la Fodor, which he is at pains to reject. (If we must have metaphors, the brain is not so much a Swiss army knife with separate blades as a woven plaid of interlinked threads.) Then six more issues get chapters of their own: categories, language, vision, motor action, problem solving, and memory (mostly external memory). Each of these chapters builds on references from the relevant literature to present an array of stimulating new insights.

In keeping with his commitment to events rather than stable states, Spivey's last chapter is not a review of what has been covered but an account of what may come next. Here, he has the mind/body problem in his sights. The present book has focused primarily on trajectories through a *neuronal* state space, but there's a bigger space on the horizon, a "fully ecological dynamic account of perception cognition and action." When dualism is finally overthrown, we will be able to see that the mind is made of "the same stuff" as the environment. Well, maybe so, maybe not. One thing is already clear: Cognitive science is on a new trajectory, and it's moving fast. Hold on to your hats!

—Ulric Neisser

This page intentionally left blank

Acknowledgments

I want each sequential change of mind in its true, knotted, clotted, viny multifariousness, with all of the colorful streamers of intelligence still taped on and flapping in the wind.
—Nicholson Baker

There are many people whom I should thank for helping me get to where I could write this book. The first people I want to thank are my family. My mother, father, and sister always tolerated and even encouraged my nerdy pursuits like computer programming and fantasy role-playing games—without which I probably would have ended up a Bohemian artist living on welfare. Steve and Sheryl Knowlton provided years of intellectual stimulation, patience, and support. I thank my wife, Melinda Tyler, for being just enough smarter than me to inspire me to work harder. Last but not least, among my family, I am grateful to little Samuel Rex Spivey for sleeping soundly in his baby sling while I write this.

The next people to thank are my intellectual family. It is perhaps egregious to actually list all the personal instructors, advisors, and colleagues whose guidance I think played key roles in developing the way I have come to think about the mind. However, when I look at the list of names, it is blatantly obvious that basically *anyone* who had this particular combination of intellectual guides (and in the particular order that I had them) would develop the viewpoint that I describe in this book. Therefore, egregious or not, their names deserve listing, as they are—in aggregate form—arguably more responsible for this book than I am. That's right—both the success and the failures of this book are more their fault than mine.

During my college years at UC Santa Cruz, particularly inspirational professors for me were Dom Massaro, Bruce Bridgeman, Ray Gibbs, and Alan Kawamoto. I also benefited from important older brother graduate students such as Brian Fisher, Ken Nemire, and Bill Farrar. During graduate school at

University of Rochester, my advisors Mike Tanenhaus and Mary Hayhoe and another older brother, Ken McRae, taught me invaluable lessons and kept me on the right track. I am also grateful to Kyunghee Koh for tricking me into falling in love with MATLAB, and to Tony Movshon for helping further my enthusiasm for computational modeling. And although I never took a course with him or collaborated with him, Jay McClelland has provided me with crucial encouragement and behind-the-scenes support in many ways and for many years.

Over the past nine years at Cornell University, I have been the lucky recipient of some incredible nurturance from my entire department, but particularly deserving of mention is the intellectual support provided by Barbara Finlay, David Field, Shimon Edelman, Ulric Neisser, and of course, the ghost of J. J. Gibson, who often walks the halls of these floors and these minds. Ulric Neisser gave me particularly helpful advice on how to make this book more encouraging and less combative (and I even managed to follow some of it). Some of the arguments in this book have also benefited from discussions with Eric Dietrich and Ken Kurtz at nearby SUNY Binghamton. Recently, I had the wonderful good fortune to serve several times on Guy Van Orden's NSF grant review panel on Perception, Action, and Cognition, where I was richly educated by the grant proposals themselves and especially by the many intense panel discussions of research and theory. I am extremely grateful to the entire panel (with its rolling membership from spring 2002 to fall 2004) and to Guy for giving me that amazing growth experience.

Essentially, I think a certain weighted combination of all of these minds, into one mind, would have written this book almost exactly as I have. And perhaps that is a valid description of what has, in fact, taken place.

I would also like to thank all of my many collaborators over the years (especially Julie Sedivy, John Trueswell, Kathy Eberhard, Viorica Marian, Daniel Richardson, and Rick Dale) whose intellectual influences induced important changes in my academic development.

This book in your hands has benefited from innumerable suggested revisions from Ulric Neisser, Daniel Richardson, Rick Dale, and many anonymous reviewers (such as Larry Barsalou, Jeff Elman, Mary Hayhoe, Art Markman, Guy Van Orden, Bob McMurray and several others whom I didn't quite manage to confidently suss out). Also, Cabot Nunlist, Jeremy Kipling, and Adam November all provided helpful early explorations into the visual search simulations in chapter 8. The incomparable Nick Hindy helped immensely with line editing and tracking down the full citations on almost 1,400 references.

I am extremely grateful to Robert and Helen Appel for their generous gift of the Appel Fellowship, which assisted greatly in providing me with time away from university duties so that a large part of this book could be written. In the summers of 2003 and 2004, at the Max Planck Institute for Psychological Research in Munich, Wolfgang Prinz and his team (Günther Knoblich, Marc Grosjean, Edmund Wascher, Peter Keller, Matthias Weigelt, Nathalie Sebanz,

and many others like Maggie Shiffrar and Bruno Repp) similarly helped me with time, money, inspiration, and Bavarian beer for working on this book. This book consumed quite a bit of all four of those precious commodities.

Finally, I wish to thank my Oxford University Press editor, Catharine Carlin, for her incredible patience and for knowing just what to say to me to trigger the necessary commitment of time and energy for writing a first book.

This page intentionally left blank

Contents

This page intentionally left blank

The Continuity of Mind

This page intentionally left blank

1

Toward a Continuity Psychology

The older dualism between sensation and idea is repeated in the current dualism of peripheral and central structures and functions; the older dualism of body and soul finds a distinct echo in the current dualism of stimulus and response.
—John Dewey (1896)

The Continuity of Mind

In an attempt to raise awareness of the benefits of emphasizing *continuous* processing, and therefore of continuous representation as well, this book ties together selected findings from neuroscience, cognitive neuroscience, cognitive psychology, ecological psychology, psycholinguistics, neural network theory, and dynamical systems theory. Without slavishly adhering to the dominant tenets of any one of those areas of research, I will build a case for a perspective on mental life in which the human mind/brain typically construes the world via partially overlapping fuzzy gray areas that are drawn out over time, a thesis that I fondly refer to as "the continuity of mind." In the service of action and communication, these continuous and often probabilistic representations are frequently collapsed into relatively discrete, rigid, nonoverlapping response categories. Each hand usually grasps only one object at a time. Each footstep is usually in only one particular direction at a time, not multiple directions. When you talk, your mouth usually utters only one sound at a time. The external discreteness of these actions and utterances is commonly misinterpreted as evidence for the internal discreteness of the mental representations that led to them. Thus, according to the continuity of mind thesis, the bottleneck that converts fuzzy, graded, probabilistic mental activity into discrete easily labeled units is *not* the transition from perception to cognition—contra cognitive psychology. Rather, that conversion does not take place until the transition from motor planning to motor execution. Everything up to and including that point is still distributed and probabilistic.

(And sometimes even the motor execution still has some multifarious gradations in it as well.)

Although this main thesis may already seem agreeable to some contemporary psychologists, not all of them may realize that it is fundamentally inconsistent with the symbolic-computation approach to cognition that traditional cognitive psychology still assumes, implicitly if not explicitly. Moreover, a wide range of other cognitive scientists, from philosophy, linguistics, and computer science, as well as other circles in psychology, have yet to seriously consider (or in some cases already strongly oppose) this perspective on the format of representation employed by the human mind. I contend that cognitive psychology's traditional information-processing approach (borrowed from the early days of computing theory), as well as certain tendencies within the more recent connectionist approach (often using strictly feedforward neural networks), place too much emphasis on easily labeled static representations that are claimed to be computed at intermittently stable periods of time. Rather than focusing on those intermittent moments when the brain's pattern of activity may be brushing up next to an identifiable *discrete* mental state representation, the continuity of mind thesis focuses on the continuous trajectory that the mind travels through the set of possible brain states—the entire *thread of thought*, if you will, rather than just the stitches that are visible on the surface of the hem.

The pattern of exposition throughout this book will be to describe a range of methodologies and findings that point to some innovative ways to observe and simulate the genuine gradedness of those mental states over time—not merely take them for granted. The continuity framework offered here draws much of its inspiration from related theoretical frameworks that preceded it, especially ecological and dynamical approaches to psychology (e.g., Gibson, 1979; Kelso, 1995; Neisser, 1976; Port, 2002; Thelen & Smith, 1996; Turvey & Carello, 1995; van Gelder, 1998; Van Orden, Holden, & Turvey, 2003). However, at the same time, this book is intended to work largely within the terminology and constraints of the dominant methodological and theoretical toolbox of contemporary cognitive psychology. For example, I will continue to use words like *representation* and *mental state*, despite their unpopularity in current dynamical and ecological approaches to cognition. However, in the process of using these traditional conceptual tools for exploring and describing the continuous nature of cognitive processing and representation, it will become clear that some new conceptual tools (and eventually a whole new toolbox) will be necessary to deal with the emerging landscape of data.

As you work your way through this book, you should expect to gradually lose some of the baggage associated with the term *representation* along the way. It need not refer to an internal mental entity that symbolizes some external object or event to an attentive central executive. Because *representation* appears unlikely to fade in use, I suggest that instead of fighting the use of the word, we can merely allow it to naturally shed that albatross of *symbolizing* something.

The word can simply continue to refer to a kind of mediating stand-in (see Markman & Dietrich, 2000), in between sensory stimulation and physical action, which is implemented largely by neuronal assemblies. However, the crucially important alteration to this stand-in function, to be touched on time and time again throughout this book, is that it is not composed of "mediating *states*" (Dietrich & Markman, 2003) but instead of something like "mediating *processes*." As the neuronal assemblies that implement most of this stand-in function never settle into truly stable states, we should not expect the mathematical description of the mediation process to settle into stable states. Therefore, my continued use of the term *representation* refers exclusively to internal mental processing that is continuous in time, is contiguous in state space, and whose function is to mediate between sensory stimulation and physical action.

The overall goal of my endeavor here is to punctuate and perturb the current instability in the metatheoretical system of cognitive science—the inconsistency between recent phenomena in the field and the accepted ways the field has for talking about phenomena in general—thereby helping enable the impending massive reorganization that the cognitive sciences so desperately need. This book is intended to map an escape route out of traditional cognitive psychology, with some hints and pointers for where to go next and build.

For those who already share this continuous, dynamical perspective on the mind, the studies described herein will hopefully provide a greater appreciation for the relationship between our multifarious, probabilistic, distributed brain states and our illusory phenomenological sense of being in one discrete unitary state of mind at a time. For those who already oppose this perspective on the mind, the many examples littered throughout this book will hopefully pose constructive challenges (some more difficult than others) for their theories to tackle. For those of you who have not already made up your minds, good for you.

These first two chapters provide a brief, easy-to-read tour through the motivation and explication of what mental representations might look like if they were indeed continuous, partially active, and partially overlapping patterns. The first thing the reader will notice is that they begin to look less like what *representation* was originally intended to mean. The reason I continue to use the term is largely to ease the intellectual transition from cognitive psychology's traditional information-processing framework to a dynamical-systems framework. I submit the notion of a trajectory through state space (a temporally drawn-out pattern of multiple "representations" being simultaneously partially active) as a replacement for the traditional notion of a *static symbolic* representation. To bring this notion to life, this chapter soon draws an analogy to the concept of a wave function in quantum mechanics, which attempts to describe the state of a system before it has been observed. Although there are explicit quantum mechanical accounts of brain states and consciousness (Goswami, 1990; Lockwood et al., 1996; Penrose, 1994; Zohar, 1995; but see Schrödinger, 1944; Scott, 1996), the continuity approach to

cognition does not depend on them. The appeal to quantum mechanics at this point is purely for expository purposes, with the goal of drawing an analogy between distributed representational brain states (that are partially consistent with multiple discrete mental states at once) and quantum mechanical superposition. Based on reactions from my colleagues, the reader will most probably either like or hate my use of this analogy. An intermediate reaction is rare.

This notion of a wave function is then connected to the way populations of neurons in the brain cooperate to represent individual perceptions. It does not seem to be the case that thoughts, ideas, concepts, categories, words, objects, or even faces are represented by solitary, individual neurons in the brain. Individual neurons appear to represent minute pieces of words, objects, and so forth. Large groups of neurons collectively represent entire words and objects. These coordinated groups of neurons are variously referred to as population codes, population vectors, cell assemblies, and cell ensembles, to name a few. For simplicity, I stick with the term *population code*. The discussion of population codes is then connected to quantitative descriptions of probabilistic representations, along with a brief treatment of the history of probability theory. After addressing the relationship between probability theory and fuzzy logic, this chapter walks the reader through two experiential demonstrations of continuous dynamical transitions through probabilistic mental states. The chapter finishes with some discussion of the conceptual reformulation that will be necessary to make sense of continuous processing and continuous representations in the mind.

The next chapter is devoted to offering some concrete (although vastly oversimplified) examples of distributed brain states and probabilistic mental states, in an attempt to make this thesis not only visualizable but indeed intuitively compelling. These examples will take us slightly (only slightly) in the direction of the conclusion favored by Churchland and Churchland (1998), that discrete nameable mental states, of the kind typically espoused by folk psychology, simply do not exist. Rather than thinking in terms of an inventory of discrete mental operands on which a central executive can perform logical operations, a continuity psychology (drawing prodigiously from ecological psychology, dynamical systems theory, and computational neuroscience) will need to think in terms of a continuous and often recurrent trajectory through a state space. Although different types of mental trajectories may be segregated into different classes for descriptive convenience, it must be recognized that the full metric range of the state space is always available to the system, in principle, and this is precisely what allows unexpected (sometimes called "productive" or "creative") organized behavior to emerge.

The third chapter reviews some concrete experimental methods that help provide a window into the continuous-time processes of the mind/brain. The fourth chapter offers some formal treatment of dynamical systems in general and describes not exactly a model but a "simulation arena" for implementing and demonstrating the complex temporal dynamics arising from biased competition (e.g., Desimone & Duncan, 1995) between idealized stable

states in a localist attractor network. Chapter 5 then outlines cognitive psychology's obsession with naming apparent discontinuities in representation and process, discusses the treatment of the overall cognitive architecture of the mind, and addresses some of the consequences that the continuous dynamical approach has for psychology. Later chapters will then review the literature, and focus on a series of experiments and idealized neural network simulations, providing compelling evidence for continuous, graded, partially overlapping representations in the mind/brain during categorization (chapter 6), language comprehension (chapter 7), visual attention (chapter 8), action (chapter 9), and reasoning (chapter 10). Finally, in the last few chapters, this book concludes by addressing some of the broader implications that a dynamical psychology has for the cognitive science notions of modularity and of representation, as well as for our own personal understandings of social interaction, consciousness, and our intellectual lives in general.

Flowing Stimulus Array, Flowing Mind

In a nutshell, the message of this book is that the human mind is constantly in motion. It does not receive individual stimuli and compute individual interpretations of them. And yet, for several decades now, the dominant frameworks of psychology have taken for granted that the mind's job is to compute individual interpretations of individual stimuli. After all, how else could we recognize what a stimulus is, if we did not activate some internal stable representation of it?

Before I get to what a temporally dynamic internal representation might be, let me first note—as J. J. Gibson (1950) did—that, in the normal everyday world, individual stimuli simply do not exist. If it is the case that individuated stimuli do not normally exist in our sensory input, then it can hardly be said that they have individuated representations devoted to them. For a given stimulus to truly be an independent entity, activating its own independent symbolic representation, it would need to be spatially and temporally separate from all other stimuli. Look around you right now. See if there are any objects that from your current perspective, are not intersecting or abutting the contours of another (potential) object. Probably not. Now move some objects around in a natural way. Take a sip from a cup, or move some paper from one place to another. As the objects move, the changes in your field of view are largely continuous through time, saccadic eye movements notwithstanding. The changes aren't freeze-frames of the object being in one location at one point in time and then suddenly in another distant location at another point in time. (Of course, it is possible to present individual objects in spatial and temporal isolation in a dark laboratory, but if that never really happens in real life, how generalizable will those lab results be?)

Now, listen to the ambient sound in your environment. Just like the visual objects abutting and occluding one another, there are several different sounds

that are overlaying one another at any one point in time. All of the sounds have a temporal duration over which they may change in complexity, pitch, volume, and so on. Just like the field of view in an interactive visual environment, the changes in your acoustic environment are largely continuous through time as well. Even the sounds that seem most "object-like," spoken words, usually abut one another in time, rarely separated from one another by even a millisecond of silence.

What this means is that the "flowing array of stimulus energy," as Gibson called it, is never presegmented into easily defined independent chunks, or stimuli—even though we feel as though we perceive it that way. Now, if the environmental stimulation impinging on our sensory systems is almost always partially overlapping in space and continuous through time, why would the mind work in a staccato fashion of entertaining one discrete stable nonoverlapping representational state for a period of time, and then instantaneously flipping to entertain a different discrete stable nonoverlapping representational state for another period of time? Why would the mind work like a computer? This book is aimed—like some other recent books (e.g., Kelso, 1995; Port & van Gelder, 1995; see also Fodor, 2000)—at responding to that question with the following answer: "It doesn't."

The New Dualism

The computer metaphor for the mind was really just the latest in a historical series of stage-based accounts of cognition. Whether the stages are the body-and-soul of dualism, or the stimulus-and-response of behaviorism, or the stimulus-and-interpretation of cognitive psychology, it may just be the idealized discrete separation of different functions that is most responsible for leading the endeavor astray. In the middle of the seventeenth century, René Descartes proposed that the mind worked by way of immaterial forces that were separate from the physical forces of our material world, and that the mind communicated with the brain via the pineal gland. Aside from the occasional personal belief in a soul, this kind of magical thinking is no longer prevalent in science. However, the same breed of dichotomous treatment of the mind as separate from the body is still quite common in the cognitive sciences—just with slightly less ethereal mechanisms being assumed.

In the middle of the twentieth century, cognitive psychology in particular, and the cognitive sciences in general, came under the spell of a new form of dualism—one fueled at least partially by our history of computing theory and artificial intelligence. Since the 1950s, when computing theory was just beginning, psychologists have likened the mind to a computer. Indeed, as other scientists have noted, humankind has made a habit of conceiving of the mind as working much like whatever happens to be the latest technological advancement. For hundreds of years, philosophers and psychologists have

written about the mind working like an hourglass, or like a clock, or like the printing press, or like a telephone switchboard, and now like a computer. Is there any reason to think this penchant for mechanical analogies is right this time?

The worrisome dualism encouraged by this mind-as-computer analogy is that it implies that the human brain is somehow functioning under very different rules, or patterns of organization, than the rest of the body and indeed, the rest of the natural world. Of course, this attitude existed well before the computer, as evidenced by Kant's (1785/1996) claim that human intelligence followed "laws, which being independent of nature, are not empirical but have their ground in reason alone." Imbuing the human brain with the power of discrete symbolic computation places it in a category by itself in nature, with all the continuous and probabilistic phenomena exhibited by the peripheral nervous system, and everything else in the natural world, placed in a different category. It becomes a "mind versus the rest of the world" attitude. But no mind is an island unto itself.

Contemporary psychology risks becoming a mockery of itself by its addiction to hypothesizing discrete discontinuities of this sort. This is precisely what Dewey (1896), from whom a quote begins this introductory chapter, was trying to curtail in his critique of the reflex arc concept. The reflex arc concept was a relatively new idea at that time, framing the questions of psychology in terms of causal arcs between (1) a sensory stimulus stage, (2) a central (mental) activity stage, and (3) an action/response stage. Essentially, studying the causal arcs between 1 and 2 *or* between 2 and 3 were to be considered legitimate scientific enterprises in and of themselves. In contrast, treating the progression of the three components as one continuous process that naturally loops back on itself was what Dewey was attempting to encourage. Actions take place over time and they continuously alter the stimulus environment, which in turn continuously alters mental activity, which is continuously expressing and revising its inclinations to action.

Behaviorism's unhelpful but long-standing solution after Dewey (1896) was to hamfistedly eliminate the second (mental) stage. After a few decades of behaviorism, the cognitive revolution, as they liked to call it, essentially resurrected that second stage and all but erased the third one (action). (At this level of description, the theoretical alteration from behaviorism to cognitivism appears minute enough that one wonders if it truly warrants being called a "revolution," see Leahey, 1992.) Essentially, cognitive psychology replaced behaviorism's emphasis on stimulus and response with an emphasis on stimulus and interpretation. These incremental adjustments to the linear treatment of the three stages reminds me of when I find myself trying to solve a toy puzzle using parametric variations of the same losing strategy, rather than trying a completely different strategy. Most of cognitive science and psychology has missed the whole point of *not* studying these stages as a linear sequence of separable components, but instead studying them as one continuous inseparable loop. Is it any wonder that our progress is plateauing once again?

Curiously, Dewey's (1896) reference to an "older dualism between sensation and idea" doesn't actually sound that old to contemporary ears. In many ways, the cognitive psychology that began with Newell, Shaw, and Simon (1958), Chomsky (1957), and Neisser (1967) among others reinvigorated the notion that sensation and perception could be part of a separate preliminary (in every sense of the word) component of mental activity, with cognition (i.e., the computation of ideas and reasoning) being a subsequent and more psychologically relevant component. Perception was just perception. But cognition was "the mind." In fact, since around the time of Neisser's (1967) *Cognitive Psychology* (see also Pylyshyn, 1984), Dewey's terms *stimulus* and *central activity* have gradually become incorporated into the central nervous system as the *discontinuous* modular suites of "perception" and "cognition". So when Dewey says, "the older dualism between sensation and idea," I have to say I feel a little bit of déjà vu.

Meet Schrödinger's Cat

Perhaps what is needed instead is a breaking down of these idealized distinctions between putative stages, a reconceptualization of mental activity as continuous in time and graded in format. To illustrate my claim that mental representations are fundamentally continuous, graded, and partially overlapping (before overt behavior converts them into discrete actions), I draw an analogy to a celebrity from popular physics: Schrödinger's cat. First, for the uninitiated, allow me to explain this feline's rise to fame. When quantum physics was gaining respectability and suggesting that the duality of light being both a wave and a particle was mathematically acceptable, there were a number of critics. Erwin Schrödinger (1935), a quantum physicist himself, became one of those critics. In his discomfort with quantum physics' claim that a particle could be simultaneously in multiple spatial locations, Schrödinger designed a thought experiment that he expected would prove quantum physics wrong. In a typical version of this thought experiment, one places a cat inside a box that also contains a chunk of mildly radioactive material, a Geiger counter, and a vial of poison gas. According to its quantum mechanical properties, this particular chunk of radioactive material is 50% likely to emit one radioactive particle per hour. If and when the Geiger counter detects this emitted radioactive particle, it triggers a device that breaks the vial of poison gas and thus kills the cat. After an hour has passed from the time you began this experiment, you might naturally conclude that there is a 50% chance that the cat is dead and a 50% chance that the cat is alive. Quantum physics would disagree with you. Quantum physics, because it allows that particle to have been emitted and not emitted at the same time, suggests that—before you look inside the box—the cat is both dead and alive.[1] Schrödinger expected the absurdity of this claim to invalidate the popular interpretation of quantum physics once and for all. How could a cat possibly be both dead and alive at the

same time?! However, to his shock and dismay, this thought experiment was *not* generally taken as proof that quantum physics must be wrong. Indeed, most quantum physicists of the time saw no absurdity in the prediction at all! As far as they were concerned, Schrödinger had beautifully demonstrated how quantum duality at the subatomic level could, under the right circumstances, be recapitulated at the macroscopic level. His cat became a popular icon for how wonderful and powerful quantum physics can be.[2]

Population Codes in the Brain

What does a confused cat have to do with the human mind/brain? The analogy I wish to draw from Schrödinger's cat to the human mind/brain is in the understanding that being *in multiple states at once* is a condition in which one can be. In fact, one might argue that it is basically impossible for the human brain to ever be in one single, entirely stable state—except for death, of course. If it were, it would not be able to gravitate out of such a state without external input. But even when the brain is cut off from all external input, during sleep or sensory deprivation, it continues to travel from one brief nearly stable state to the next: we dream, or we hallucinate, or we experience a "stream of consciousness."

When we look at how the brain encodes information, we see that it is a lot like the wave function that characterizes the multifarious state Schrödinger's cat is in. The majority of neurons studied in mammalian brains send their signals in the form of relatively discrete all-or-none action potentials, brief but intense depolarizations (1–10 milliseconds) of their electrochemical membrane potentials. However, it does not appear to be the case that the firing of individual neurons is used to signal the presence of things like objects, words, and concepts (see Damasio & Damasio, 1994; Hebb, 1949; Pouget, Dayan, & Zemel, 2000; Rose, 1996; see also Barlow, 1972). For some time now, neuroscientists have been able to record the activity of many neurons at once in various regions of the nonhuman primate brain and have generally been finding that *populations of neurons* participate together to embody a representation. For example, in the 1970s, David Sparks and colleagues showed that the neural signal that tells the eye muscles to move the eyes in a particular direction is made up of many neurons, in the superior colliculus of the macaque monkey, each of which represents a different direction of eye movement. It is the *distribution of activity* across this population of neurons that determines the direction of the eye movement, not just the activation of those neurons that specifically code for the actual direction the eyes wind up going in (Sparks, Holland, & Guthrie, 1976). In the 1980s, Georgopoulos and colleagues found similar evidence for population codes of arm movements in the motor cortex of the macaque (Georgopoulos et al., 1982). Moreover, it appears that population codes are used not only for representing and producing motor *output* (e.g., eye and arm movements) but also for representing perceptual *input*.

For example, in the 1990s, Wilson and McNaughton (1993) demonstrated that ensembles of cells in the rat hippocampus cooperate to encode the animal's knowledge of what environment it is in. And Tanaka (1996, 1997) showed that visual objects (faces included; see Gauthier & Lokothetis, 2000; Perret, Oram, & Ashbridge, 1998) are represented by populations of cells within the inferotemporal region of visual cortex in the macaque.[3]

One of the things that makes population codes (i.e., distributed representations) robust and powerful is that under noisy or degraded stimulus conditions or following physical injury, they will often still be able to approximate the original input signal: *graceful degradation* (Rumelhart & McClelland, 1986a). For example, imagine that a particular set of 100 neurons participate in the representation of your grandmother's face, such that when you look at her, the ideal, perfect recognition would happen if those 100 neurons were at their appropriate activation levels (firing rates). If she laughs and covers her mouth, then some of those 100 neurons will reduce in activation because the parts of her face to which they especially respond are occluded. Nonetheless, if 80 of those 100 neurons are still doing what they are supposed to do, that population code for grandmother (with its 80% "confidence") will still be by far the most coherent code available in the brain. In contrast, if you had only one neuron devoted to recognizing grandmother, this "grandmother cell" (Lettvin, 1995) may not be able to do its job when grandmother covers her mouth, turns her head, or makes a funny face. You'd suddenly fail to recognize her!

What this means is that with population codes, we are *always* dealing with internal representations that have what you might call percentages of confidence (or probabilities, loosely) associated with them. The image on your retina of your grandmother will almost never be the same at any two points in time. Therefore, the input to those 100 neurons (your grandmother population code) will never be exactly perfect to turn them all on. This population code will be in a nearly stable state. What often happens then is that the connections between the members of this population code will pass the activity back and forth and increase the percentage of them that are active. This *pattern completion* process (e.g., Grossberg, 1980) will gradually increase the population code's "confidence," and thus its probability of producing an associated behavior—such as pushing air out of your lungs to vibrate your vocal chords while articulating parts of your mouth to make the sound, "Grandma!" Importantly, that discrete behavior—saying one particular word and not any other words—is often interpreted by the people around you as indicating that your internal representation for grandmother is 100% "confident." The continuity of mind thesis posits that your representation is *not* 100% confident and can *never* be 100% confident.

Although the process of pattern completion will increase the total activation (or probability) of a representation over time, its associated action will be produced long before the representation ever reaches maximum activation (or probability 1.0). This action (even something as benign as moving your eyes to a chair, near Grandma, that you plan to sit in) then inevitably changes

the sensory array, so that the original input to that population code is now crucially altered, and a new pattern completion process must begin—gravitating the system toward a new and different probabilistic mental representation.

Versions of Probability

If we accept this account of population codes as probabilistic representations of multiple unitary concepts (see Zemel, Dayan, & Pouget, 1998), for example, 0.8 Grandma, 0.02 Kathryn Hepburn, 0.01 Mother Teresa, and hundreds of other representations with very low confidence, that together add up to 1.0, then we begin to see how the mind is indeed like Schrödinger's cat: in multiple identifiable states at once. However, we must acknowledge that this is using a particular connotation of *probability*, a term which has taken on many senses in the last couple of centuries. Because a form of probabilism is infused in a great deal of the theoretical treatment throughout this book, the following section will describe some of the different interpretations of probability, cover some of its history, and also jog your memory with just a touch of math.

In the eighteenth and nineteenth centuries, a great many philosophers, mathematicians, economists, and physicists (as well life insurance statisticians!) were employing the tools of probability to essentially make predictions about future events. Much of early probability theory was actually developed in the interest of using death statistics (i.e., mortality tables) to determine profitable life insurance coverage and premiums. Crucially, the dominant meaning of probability at the time was one of describing the likelihood (as a value between 0 and 1) that a future event will end up discretely in one state or another. Thomas Bayes formulated an extremely influential theorem that instructs exactly how to do this (Bayes, 1763/1958).

Let's walk though an example. Imagine that you just lost all your money at the roulette table of a new casino. Let's assume you usually at least break even at roulette (95% of the time), so you're now suspicious—for the first time in your life—that the wheel might be rigged. Bayes's theorem lets you pit the likelihood of your rare event against the general likelihood of casinos cheating, to calculate the probability that this particular casino just cheated you. For the sake of argument, assume that based on crime reports, 1 out of 100 casinos rig their roulette tables to cheat gamblers out of their money. Understanding equation (1) is easier than you might think.

$$\mathrm{P}(C \mid L) = \frac{\mathrm{P}(C)\,\mathrm{P}(L \mid C)}{\mathrm{P}(C)\,\mathrm{P}(L \mid C) + \mathrm{P}(notC)\,\mathrm{P}(L \mid notC)}. \tag{1.1}$$

Let $\mathrm{P}(C \mid L)$ be read as "the probability of this casino cheating, C, given that you just lost all your money, L." For the numerator, we multiply the base rate,

or prior probability, of C (i.e., 1/100) by the probability of your losing if the casino cheated, $P(L \mid C)$; let's assume that would be 1.0. In the denominator, that same product, $P(C)\ P(L \mid C)$, must be added to the probability of the casino being fair, $P(notC)$, multiplied by the probability of your losing at a fair casino, $P(L \mid notC)$. This is necessary to normalize your suspicion against the alternative possibility: that you just got unlucky. Dividing the numerator ($0.01 * 1$) by the denominator ($0.01 * 1 + 0.99 * 0.05$), results in $P(C \mid L) = 0.168$. Certainly a much higher likelihood than the base rate of 1 in 100, but not quite enough confidence to warrant contacting the police. Perhaps if it happens to you three times in a row at that same casino, then it might be time for an investigation . . . or then again, maybe you've just lost your touch.

Probability theory also allows us to compute the probability of *combinations* of events. For example, the probability of a flipped coin coming up heads twice in a row is computed by simply *multiplying the probability of the first event with the probability of the second event*: $0.5 * 0.5 = 0.25$. (Of course, this only really works when the probabilities are independent of one another.) The probability of that casino *not* cheating, even though you've lost at roulette three times in a row there, could be calculated as $(1 - 0.168) * (1 - 0.168) * (1 - 0.168) = 0.576$. Thus, it would appear that Bayesian theorists can make some pretty sophisticated predictions, not only of individual events but also of combined events.

However, the Bayesian interpretation of those mathematical results is not accepted by everyone. A frequentist's view of probability would emphasize that although the 0.25 probability of flipping two heads in a pair of coin flips tells us to expect about 25 heads-heads out of 100 pairs of coin flips, probability can say nothing about which face of the coin is actually up on any one flip. We must rely on observation to tell us that. In the strict frequentist account of probability, there is no discussion of the degrees to which an individual event is *likely* to be in one state or another—and certainly no acknowledgment of the degrees to which an individual event is *in one state and another at the same time*!

The way I would like to encourage the reader to think of probability in the mind is a far cry from the frequentist's interpretation and even subtly different from the Bayesian interpretation. The continuity of mind thesis holds that simultaneously partially active mental representations can be treated as summing to 1.0 and thus may represent the probability of their individual associated actions being elicited. In this view, it is the fact that the body's effectors (limbs, hands, eyes, speech apparatus, etc.) can each typically only do one action at a time, which causes the multifarious amalgam of mental states to warp itself over time toward largely approximating only one mental state just long enough to produce that mental state's associated action. Thus, when relating these multiple graded mental states to possible actions, the thesis looks decidedly probabilistic, but when examining the mental states for their own sake, the thesis might be best compared to fuzzy logic.

Following some initial work by logicians on elements of a formal logic that allowed for "vague" truth values, Lotfi Zadeh introduced the notion of fuzzy logic (Zadeh, 1975; see also Massaro, 1997). In fuzzy logic, the truth value of a proposition (such as "Donald is rich") has a range between 0 and 1. Moreover, the truth value of a conjunction of propositions (such as "Donald is rich and I am poor") is equal to *the truth value of one proposition multiplied by the truth value of the other proposition.* Sound familiar? The mathematics of fuzzy logic and the mathematics of probability are essentially the same. It is the interpretation that differs. Fuzzy logic takes the mathematical results of traditional probability statistics and accepts them at face value as "the (multifarious) state of the system," not as "a prediction of the possible discrete states the system might be in." This is precisely what quantum physics does with its mathematical description of the probability that Schrödinger's cat is dead and the probability that it is alive. It accepts the math as a *conjunctive description* of the world, not as a *disjunctive prediction* about it.

"Warping" the Probabilities

You can begin to see the tension here between the notions of probability and fuzzy logic. I will perhaps add to that tension when I note here that the "probabilistic" activations of mental representations discussed throughout this book often do not adhere to the mathematics of Bayesian probability theory (see chapter 4 for details). From this perspective, my use of the term *probability* may seem somewhat glib. The conjunctive description of mental contents provided by fuzzy logic is converted into a disjunctive prediction, via probabilities, of the motor responses being recorded by the psychological experimenter. The way in which *probability* truly does apply here is in the stipulation that these fuzzy logical activations of mental states are treated as "the probability that the mind will activate a motor action that is associated with a particular perceptual category." However, because their activations change continuously, these partially active mental representations should not really be interpreted as "the mind computing the probability of a given stimulus belonging to a particular category." At a very deep level, this claim is actually quite shocking, if not preposterous. It amounts to saying that A and B (below) are true, but C is *not always true.*

A. There are Bayesian probabilistic relationships between external states in the environment.
B. There are Bayesian probabilistic relationships between mental states in the mind and motor actions in that environment.
*C. There are Bayesian probabilistic relationships between external states in the environment and mental states in the mind.

What could be so special about that transition from stimulus to percept (statement C) that it dares defy the mathematics of Bayesian probability?

In fact, a considerable amount of research in a subfield that calls itself Bayesian perception adheres rather strongly to statement C (e.g., Kersten, 1991; Knill, 1998; see also Rao, Olshausen, & Lewicki, 2002). Bayesian approaches to perception usually acknowledge the gradedness of internal mental states; however, they still tend to treat them as static in time. The temporal dynamics of cognition is largely ignored by the Bayesian approach to perception. Thus, although an experiment in Bayesian perception can often demonstrate an accurate mathematical prediction (in the form of some probabilities) about the overt categories into which an observer will place her percepts, it usually demonstrates nothing about the temporally extended process by which the sensory input eventually led to a particular categorical response. In the context of having considered the pattern completion process exhibited by neural population codes and by attractor dynamics, this two-step process of *stimulus* and then *probability* is reminiscent of the two-step "stimulus and then response" attitude criticized by Dewey (1896).

There are properties inherent to dynamical systems that are often responsible for the mind not quite adhering to probability theory. There is a kind of momentum that the mind develops as it travels through the state space, causing it to warp and exaggerate its deterministic influences. The mind has a tendency to gravitate closer to the nearest attractor (mental state) than warranted. That is, dynamical systems often settle toward stable states, with one attractor being almost, but not perfectly, satisfied (i.e., its "interpretation" of the input being somewhere near 1.0 probability)—even when the input is unresolvably ambiguous. As mentioned earlier, this pattern completion process takes place over a period of time (whether it be a few hundred milliseconds or a few seconds). One must look inside this pattern completion process to find evidence of probabilistic mental states. Too often, researchers examine the final result of a mental process, such as the category or accuracy of the solicited overt motor response. Although informative for characterizing the hypothesized representations that putatively get computed, this mindset largely neglects the process of settling toward those representations and the fact that many amalgams of representations are often considered along the way. The continuity of mind thesis is not particularly aimed at discounting the expository usefulness of those idealized discrete representations of pure mental states. Rather, it is aimed at bringing to the reader's attention the fact that "getting there is half the fun."

Nonlinear Attraction, Stability, and Instability in Visual Perception

Figure 1.1 shows a cartoon example of a two-dimensional perspective on a vector landscape for the high-dimensional state space of a dynamical system. This is a way to visualize the temporal dynamics of a system's state as it would traverse through its state space. Pick a location anywhere on that

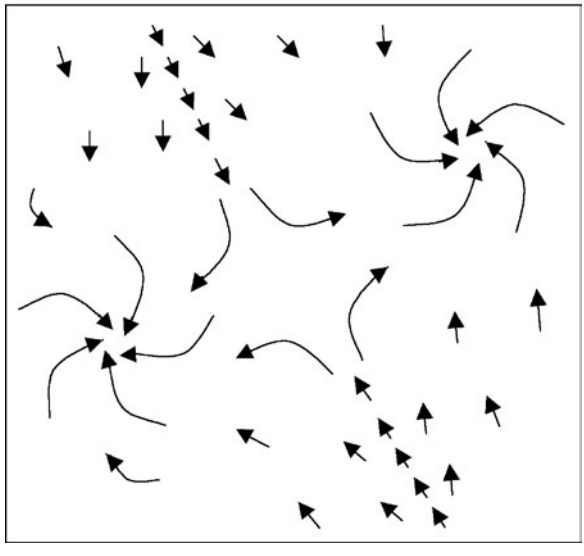

Figure 1.1. A schematic example of a vector landscape for a dynamical system with two attractor basins.

two-dimensional map (recognizing that it would actually correspond to a location in the high-dimensional state space of the dynamical system itself), and put your finger on the location. There are arrows nearby that (with a little interpolation) give an indication of what direction the system would move in. Longer arrows imply stronger attraction and hence faster movement. Move your finger in the direction of the attraction, and check the direction of the arrows near your finger's new location. Continue moving your finger so, and you'll simulate the continuous trajectory of a dynamical system as it moves through its state space. Note that the two attractor basins are spiral-shaped, such that the system would take a while to settle motionlessly into the point attractor, tending to make smaller and smaller orbits almost indefinitely. Thus the vector landscape itself is likely to change shape (due to new sensory input and/or planned motor output) before the state of the system actually becomes static.

Figure 1.2 shows a different kind of rendition of a similar state space manifold. The energy landscape in figure 1.2 shows the two attractor basins as actual bowls in the surface. The vertical axis is treated as energy, and the dynamics will always push the state of the system toward a reduction in energy. Imagine placing a marble on the mesh surface of figure 1.2, and envision where it would roll. Thus would be the trajectory of the system over time.

Any time there is more than one attractor in a dynamical system, it is considered a *nonlinear* dynamical system. With more attractors comes greater potential for any given trajectory to meander quite nonlinearly in its high-dimensional state space. What is crucial to defining a dynamical system is its

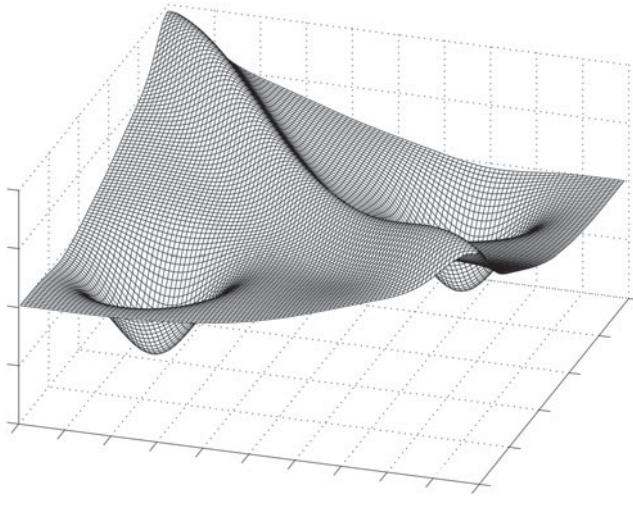

Figure 1.2. An energy landscape similar to the vector landscape in figure 1.1.

balance of stability and instability (e.g., Glendinning, 1994; Spencer & Schöner, 2003; Ward, 2002; see also Bak, 1994).[4] Nonlinear attraction is how a system achieves relative stability, as it travels from unstable point to unstable point in state space to gradually settle into the basin of a point attractor. However, too much stability can be a bad thing. If the system settles all the way into the point attractor—rather than just orbiting its basin[5]—then the system is stuck there until external perturbation dislodges it. In thermodynamics, this kind of true stability is affectionately referred to as heat death.

One easy way to undo a relatively stable state in a dynamical neural system, and reachieve instability, is through fatigue. If a neural population code is continuously stimulated for a significant amount of time, one can naturally expect that the refractory periods of the individual neurons will accumulate in number and duration until it becomes quite difficult to substantially excite that population code for some time. This has been demonstrated in neural firings rates in monkeys (e.g., Baylis & Rolls, 1987; Carandini, 2000; Maffei, Fiorentini, & Bisti, 1973; Sekuler & Pantle, 1967), in human neuroimaging (e.g., Noguchi, Inui, & Kakigi, 2004; Thompson-Schill, D'Esposito, & Kan, 1999), and in neural network simulations (e.g., Huber & O'Reilly, 2003; Kawamoto & Anderson, 1985). This fatigue of the population code results in the reduction of its attraction strength in the state space, and other nearby attractors (population codes) will now be able to pull the system toward them. Such neural fatigue is a common explanation for a wide range of perceptual alternations and illusions, including the following experiential demonstration. It has long been suggested that the perspective alternations of the Necker cube (figure 1.3) are due to fatigue, or satiation, of neural representations (e.g., Orbach, Ehrlich, & Heath, 1963; see also Köhler & Wallach, 1944).

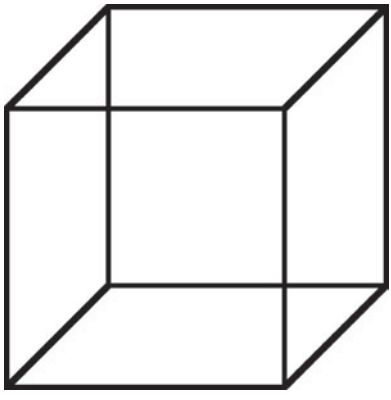

Figure 1.3. The Necker cube. At first glance, it appears to be a wireframe box with one particular perspective, for example, viewed from slightly above it. However, after staring at it for a few seconds, the perspective will change to one in which the box is being viewed from slightly underneath it. See text for discussion of these perspective reversals.

When looking at this wire frame cube, the lower square will often appear to be the front (or closer) panel of the cube, as if your head is slightly above the cube and you are looking down at it. However, after staring at it for several seconds, your percept will switch to having the upper square appear to be the front panel, as if your head is slightly below the cube and you are looking up at it. A few seconds later, the percept will switch back for a little while. As the perspective with the upper square appearing in front is a somewhat unusual one (requiring the cube to be suspended in air or resting on a glass shelf), it is perhaps not surprising that this percept usually lasts for a slightly shorter period than the more canonical one (see Wallach & Slaughter, 1988). Over time, this oscillation between perspectives of the Necker cube tends to increase in rate. Thus, if you were to report when the perspective reverses over time, the graph of those reversals would look something like figure 1.4.

The bistable pattern of Necker cube perspectives has been described as a dynamical system in which two attractors compete against one another

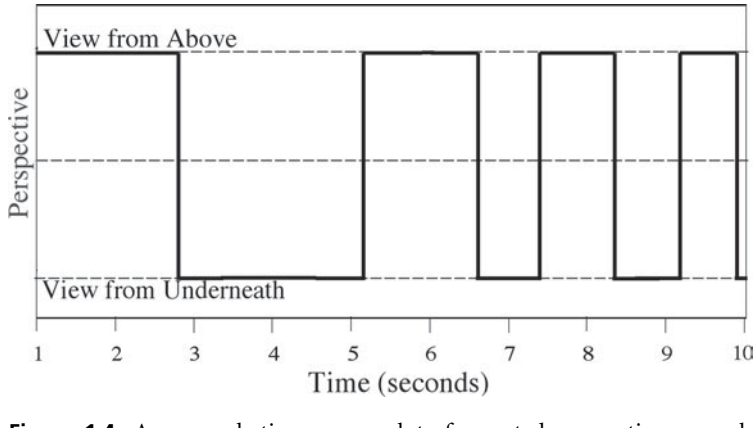

Figure 1.4. An example time course plot of reported perspective reversals during viewing of the Necker cube.

(DeMaris, 2000; Kawamoto & Anderson, 1985; Kelso, 1995; see also Hock, Kelso, & Schöner, 1993, and van Leeuwen, Steyvers, & Nooter, 1997, for similar dynamical treatment of bistable visual input). The perceptual alternations observed with the Necker cube (as well as other ambiguous figures, such as the classic vase/faces silhouette and the Schröder stairs) are consistent with a dynamical systems account of a nonlinear trajectory settling into one attractor basin and then into the other, and back, and so on. However, flipping back and forth between two relatively stable states is something that a logical symbolic (computerlike) system can do as well. What a logical symbolic system cannot do is visit intermediate gradations between the two identifiable states, as a dynamical system naturally does. Therefore, the important observation to note regarding the perceptual alternations of the Necker cube is not simply that they bounce back and forth but that they take a nonzero amount of time to do so. The transition from one identifiable percept to the other is not instantaneous. Based on numerous informal phenomenological reports, when a stable Necker cube perspective begins to transition to the alternative perspective, it seems to take somewhere around half a second for that current percept to finally give way and be replaced by the alternative percept. If this is the case, then the actual perceptual state is not quite accurately described by the instantaneous transitions plotted in figure 1.4. The discrete step-function quality of the data may be more an artifact of the constraints of the experimental task, for example, "press this button or that one, not both," than a true indication of the internal mental state of the observer. (For similar circumstances of *response discreteness* being misinterpreted as *mental discreteness*, see the discussion of categorical perception in chapter 6.) Rather than discretely jumping from one perspective to the next with a step function, perhaps it would be more accurate to plot the Necker cube perspectives as transitioning with a sigmoid function (i.e., an S-shaped curve). See figure 1.5.

In fact, some observers report being able to perceive some visual properties of the intermediate conditions *during the transition.* The perceptual transition is often described as the back panel moving closer in depth and the front panel moving away in depth, until they are at the same depth plane, and the image looks something like a wire frame mobile that is collapsed. The two panels continue their movement, crossing each other, and eventually take each other's previous places. And, believe it or not, there is even one introspective report of the percept "getting stuck" in one of those intermediate conditions for a couple of seconds!

This account is based on introspective reports, of course, and therefore should be taken with a grain of salt. But then, so is the original measure of the Necker cube's perspective reversals, as exemplified in figure 1.4. The only difference is that the introspective report for the data in figure 1.4 is methodologically constrained to a two-alternative forced choice. That is, the observer is explicitly instructed to press one button when one perspective comes into view, and then press another button when the other perspective comes into view. Pressing both buttons at once is not an option. This requirement of

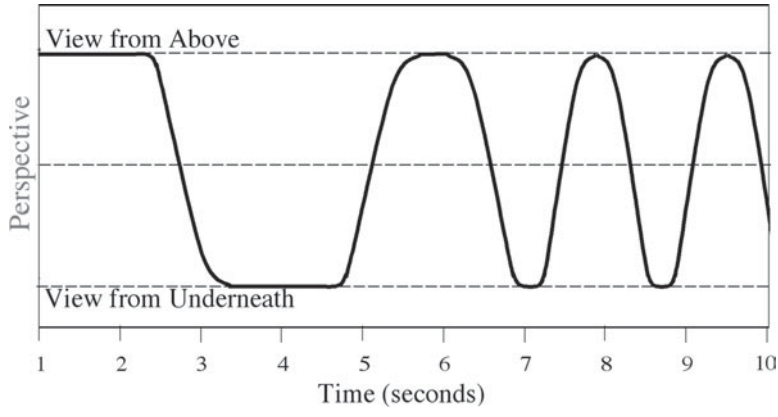

Figure 1.5. A hypothetical time course plot of the actual perceptual state during viewing of the Necker cube. The flat horizontal portions of this oscillating curve are, in dynamical systems terminology, the stable states, where the system is nestled in one of the attractor basins in the state space. The diagonal and curved portions characterize the periods of time when the system is unstable and not inside either attractor basin, but is in the process of being attracted to one of them.

discrete, categorical responses is quite common in cognitive psychology. In contrast, if we allow observers to (at least attempt to) provide more than just a selection of one of two categories, then we have a chance at obtaining a measure of the continuous probabilistic character of mental activity. Throughout this book, there are many different examples of ways to measure and observe, with considerable experimental rigor, that continuous probabilistic character of mind. Consider the sigmoid curves in figure 1.5 our first data visualization (of many to come) of what I call the continuity of mind.

Another compelling data visualization of the continuous manner in which a percept gradually comes into view can be found in neurophysiology research. Recordings from multiple neurons in the inferotemporal cortex of the macaque monkey suggest that it takes a few hundred milliseconds for the right population of cells to achieve their appropriate firing rates for fully identifying a fixated object or face (Rolls & Tovee, 1995; see also Perrett, Oram, & Ashbridge, 1998). The cumulative information (in bits) provided by an inferotemporal neuron in the service of recognizing a face or object accrues continuously (though nonlinearly) over the course of about 350+ milliseconds (see figure 1.6). About 80 milliseconds after the presentation of the visual stimulus, these cells begin firing, and during the first 70 milliseconds of firing, about 50% of the total information to be encoded is already accumulated. Thus, very quickly the network is able to project itself into the right general "neighborhood" in its state space. (This allows some coarse visual discriminations to actually be made with 100 milliseconds or less of stimulus presentation

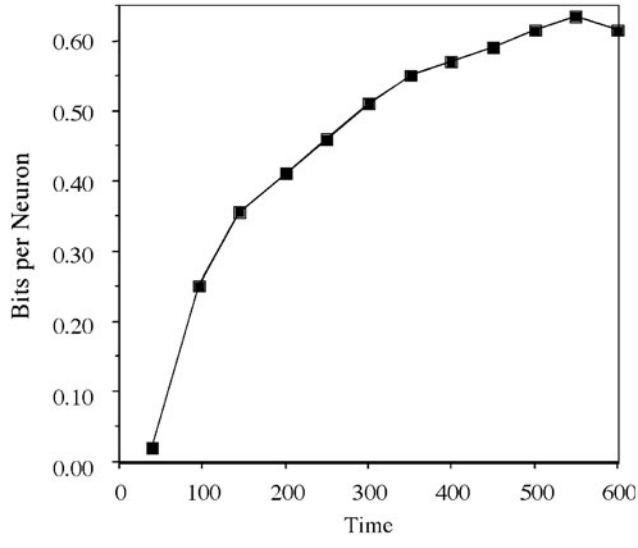

Figure 1.6. Average cumulative information accrued over milliseconds by inferotemporal cells representing objects and faces (adapted from Rolls & Tovee, 1995).

time; see Potter, 1976, 1993; Van Rullen & Thorpe, 2001.) However, over the next 200+ milliseconds, the process of object or face recognition is still *in progress*, during which the remaining 50% of the information to be represented by the distributed population code is gradually accumulated.

Admittedly, 350 milliseconds for a population code to be in transit on the way toward achieving its potentially stable state might not seem like a lot of time. The stable states depicted for the Necker cube in figure 1.5 certainly take up a substantial amount of the total time. Are the transition periods perhaps just interesting curiosities, and the important observation is that a stable state is eventually reached, and it is *that* on which logical mental computations are performed? I think not. Throughout the course of this book, I hope to convince you that the transitions are the important observations, not the seemingly stable states. It is my hypothesis that in more complex visual (as well as auditory, olfactory, etc.) environments, the proportion of time spent in these unstable regions of state space—that is, in the process of traveling toward an attractor basin, but not in one yet—is actually much greater than the proportion of time spent in relatively stable (or, more precisely, metastable) orbit-prone regions of state space.

This gradual accrual of the information comprising a population code (figure 1.6) has powerful consequences for how we conceptualize what the brain is doing when we go about our business of naturally perceiving the world around us. Consider how your eyes move around a complex scene like the

one in front of you right now. Your eyes rest, with the two foveas fixating a particular location in the visual field, for about 200–300 milliseconds on average (e.g., Rayner, 1998). They then make a fast, ballistic jump (lasting a few dozen milliseconds or so) away from that location to fixate another location in the visual field. After resting there for another 200–300 milliseconds, they jump yet again to another location. Each new fixation brings a new word, object, or object part, into the high-resolution view of your foveas for little more than a quarter of a second. Now, if it takes almost half a second for the appropriate population code to get fully settled in recognizing a fixated object, but your eyes normally move to a new object every quarter of a second, how can the brain achieve a genuinely stable state for any object recognition event?

Perhaps a stable state is not necessary. Perhaps the relevant neural networks in the brain need only approach an attractor basin in their state space closely enough so that it is unambiguously the most coherent of the many partially active population codes, and then that attractor's associated motor actions and anticipated perceptions go on to carry out their own activation processes. From this perspective, the image of a mental trajectory is now decidedly different from one in which the state of the system lands in one attractor in state space, to consider one thought or percept, and then it lands in another attractor to consider another thought or percept. Rather, the image is one in which the neural system continuously traverses intermediate regions of its state space and occasionally briefly brushes up near an attractor basin just long enough to bring that attractor's associated percepts and actions into prominence. The emphasis is on the journey, not the destinations.

Thinking of objects (or words) as living in a high-dimensional space is a little bit like shooting pool, if you treat the cue ball as the current state of the system, and the object ball (the one you're aiming at) as the next upcoming attractor. A good pool player thinks not only about how to sink the object ball but also about where the cue ball will go after that. Where the state of the system goes after brushing up next to the current attractor is incredibly important. The process of recognizing the *next* word or object does not begin from some neutral central location in state space. It begins from where the system last left off. In a dynamical neural system, the mind travels a continuous trajectory in this state space; it cannot teleport itself to neutral locations in the state space in between recognition events, the way a computer can instantaneously flip its states to some context-free unbiased baseline. Therefore, precisely where in state space the previous word/object left the system has a powerful influence on the trajectory it takes to get to the location in state space corresponding to recognition of the next word/object. Hence, one should expect "priming" effects from the previous word/object on the recognition of the current word/object. And of course, as every cognitive psychologist knows, the literature is rife with reports of words priming one another (e.g., Lukatela, Lukatela, & Turvey, 1993; Neely, 1977; see also Trueswell & Kim, 1998) and reports of objects priming one another (e.g., Cooper, Beiderman, & Hummel, 1992; Gauthier & Tarr, 1997; see also Dill & Edelman, 2001).

Nonlinear Attraction, Stability, and Instability in Language Processing

If you are one those people who feel as though they can catch a glimpse of what the Necker cube looks like—sort of—during the time course of its transition from one perspective to the other, then you have witnessed, firsthand, the continuity of mind. However, if such a glimpse eludes you, fear not. I have a second experiential demonstration of the neural fatigue of a population code that just might work for you. In much the same way that staring at a bistable visual image and perceiving it in one of its two possible perspectives for several seconds essentially overexposes the population of neurons that represents that percept, one can induce the same kind of effect in language. Look at the word in figure 1.7. This is a familiar, easy-to-recognize word. On looking at it, you feel as though your mind achieves a stable interpretation of its meaning. However, if you overexpose the system to this input, you can actually fatigue that meaning to the point that it no longer produces a stable state but instead a clearly introspectively unstable one. Fixate the word in figure 1.7 and read it out loud to yourself, about once per second, for one minute. Each time you say the word, run a kind of mental inventory check on what the word is making you think of at that point in time.

For most people, most of the time, the meaning of the word seems to disappear after many repetitions. The word will begin to look and sound like an unfamiliar nonsense word or perhaps a word from a foreign language. Sometimes you can notice the gradualness with which the original meaning fades. Moreover, one can also occasionally become aware of strange associations that arise, which are indicative of more than just a loss of the original meaning but instead a gradual transition of the system into unusual regions of state space. That is, as the neurons comprising the population code for the meaning of giraffe begin to fatigue, other slightly related populations codes become relatively more prominent. For example, as the meaning "a very-long-necked orange quadruped from Africa" dwindles, you might find yourself making peculiar observations, such as the fact that the *g* is ambiguous with respect to its pronunciation (e.g., as in *giant* and *gimlet*). Or similar sounding words may come to mind, such as *raffle*, *draft*, or even *rafter* (if you speak fast and the syllables exchange order). Or perhaps, you'll think of names, like

Figure 1.7. To demonstrate semantic satiation, look at this word and read it out loud to yourself, about once per second, for a minute. As the repetitions continue, the meaning of the word will seem to fade.

Al Jaffe, a cartoonist for *Mad* magazine, or Daniel Jurafsky, a well-respected computational linguist and recent MacArthur Fellow. One colleague even said that the word began to sound like a pretentious French-derived adjective, as in "he's *so* giraffe," meaning something like *gauche* or *jejune*. This odd stream of consciousness, occurring as the original meaning diminishes, should not be surprising if one conceives of word meanings as living in a high-dimensional state space. With each dimension being represented by the activation of its corresponding neuron in the network, reducing the coherence of the population code for the word *giraffe* unavoidably means increasing the coherence of other population codes in nearby regions of state space. As the system gravitates away from the *giraffe* attractor basin, it cannot help but travel somewhat near others. Figure 1.8 is a simplified caricature of a hypothetical two-dimensional perspective through this high-dimensional space that would allow one to watch the trajectory of the system exhibiting fatigue of the *giraffe* attractor and therefore meandering slightly near some other attractors.

This bizarre phenomenon has actually been well studied for decades and is commonly referred to as semantic satiation (e.g., Jakobovits, 1967; Smith & Klein, 1990; see also Tuller, Ding, & Kelso, 1997). Although early theories about semantic satiation treated the effect as though it was a discrete loss of meaning that took place at a particular point in time (e.g., Mason, 1941; Severance & Washburn, 1907), Lambert and Jakobovits (1960) demonstrated

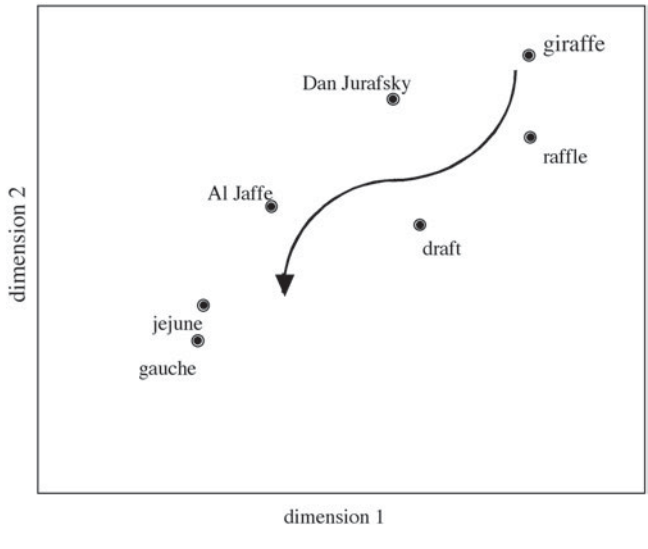

Figure 1.8. During semantic satiation, the meaning of a word diminishes, and similar associations can come to mind. This schematic two-dimensional state-space depicts a hypothetical trajectory away from the satiated word, *giraffe*, and skimming near other words/concepts in the space.

the gradual nature of this reduction in meaning over time. Using Osgood, Suci, and Tannenbaum's (1957) semantic differential measure, which projects the meaning of a word into a several-dimensional space, Lambert and Jakobovits had participants provide responses for locating the word in that space after longer periods of word repetition. As semantic satiation accrued over more repetitions, the resulting projections of word meanings in the semantic differential space indicated a gradual and continuous movement toward but not all the way into the null origin of the space.

Osgood et al.'s (1957) three to six dimensions for representing the meanings of words was an important breakthrough, but it was still quite different from the high-dimensional state space of a neural network. Their dimensions were based on rather abstract concepts, such as good/bad, active/passive, and potent/impotent, for which participants simply provided metacognitive ratings for any one word (e.g., on a scale of $+3$ to -3, how good/bad, active/passive, and potent/impotent is a giraffe?). Moreover, the physical mechanisms by which these abstract dimensions might be instantiated were not forthcoming. In fact, precisely because the actual space in which these words live is high-dimensional, which is merely approximated by Osgood et al.'s abstract dimensions, *almost any* set of concepts that are sufficiently different from one another could probably serve as the basis vectors for a several-dimensional projection of that high-dimensional neural space (e.g., Edelman, 1998, 1999). For example, if one had participants report how similar any word is to a peanut, an airplane, and a horse, one could probably produce a three-dimensional mock-up that would exhibit important clusterings of abstract concepts such animate/ inanimate, natural/artifact, and so on. But it's probably not the case that the principal dimensions on which our brains encode the world are *peanutness*, *airplaneness*, and *horseness*.

Nonetheless, Osgood et al.'s (1957) insight that word representation should be carried out in a metric space, where graded similarity is easily embodied as the distance between representations, was important—yet was quickly swept under the rug as the computer metaphor of the mind took hold in the 1960s. In cognitive psychology, the dominant account of word representation became symbolic entries for words (like in a dictionary), with their relationships to one another encoded by logical rules and/or sharing of an integral number of discrete semantic features. Essentially, if one could easily imagine coding the representation scheme in the popular programming language of the time (LISP), then it was considered a legitimate representation scheme. Coding a high-dimensional metric space, with each word being a continuous vector in that space, was not what LISP was best at doing. However, now that symbolic programming is nowhere near as dominant as it was in the 1960s and 1970s, and numerical computation has become quite popular, perhaps it is not surprising that high-dimensional geometric accounts of word representation are becoming accepted again (e.g., Landauer & Dumais, 1997; Lund & Burgess, 1996; Schutze, 1993).

Deprogramming the Cognitive Psychologist

The change in styles of programming languages from symbolic to numerical is only one of many transitions that have recently taken place to help set the stage for what promises to be the next paradigm shift in psychology and the cognitive sciences. For example, connectionism, though not quite becoming the dominant paradigm in psychology, managed to make the concept of distributed representations an acceptable notion (e.g., Clark, 1993; Elman et al., 1996; O'Reilly, Munakata, & McClelland, 2000; Rogers & McClelland, 2004; Rumelhart & McClelland, 1986a; but see Dietrich & Markman, 2003; Fodor & Pylyshyn, 1995; Marcus, 2001). One could argue that much of the connectionist literature has devoted slightly too much of its attention to trajectories through synaptic-weight-space as an account of learning and not enough to trajectories through activation-space as an account of real-time processing. Nonetheless, the step to having knowledge live as partially overlapping distributed representations in the high-dimensional state space of a network has been a crucial departure from cognitive psychology's traditional symbolic computation approach.

Moreover, improvements in continuous and semi-continuous measures of cognitive processing have helped open the door to visualizing the continuous dynamics of mental activity. For example, speech shadowing (repeating continuous speech as quickly and accurately as possible) provided important insights into language processing (e.g., Marslen-Wilson, 1973, 1975). Recordings of electrical potentials from the scalp (e.g., Hillyard & Kutas, 1983) as well as from the peripheral muscles (e.g., Tuller, Kelso, & Harris, 1982) have provided continuous measures of a wide range of perceptual, cognitive, and motor processes. Recording from multiple neurons at once (e.g., Georgopoulos et al., 1982), recording from neurons in awake behaving animals (e.g., Motter, 1993), and microstimulating neurons in awake behaving animals (e.g., Gold & Shadlen, 2000) has provided concrete examples of the distributed probabilistic states in which neural systems spend much of their time. Eye tracking has provided real-time semi-continuous measures of language and vision (e.g., Rayner, 1998; Tanenhaus et al., 1995). These relatively recent advancements in methodologies (as well as many others; see chapter 3) have made it possible to catch glimpses of the graded states that the mind travels through on its way to produce discrete actions.

Another development in the cognitive and neural sciences that assists in placing us at the brink of a significant movement away from traditional cognitive psychology is that of dynamical systems theory. As a field of its own, dynamical systems theory has advanced a great deal in both sophistication as well as popularity since the days of Hamilton, Boltzmann, and Poincare. For example, recent treatments of dynamical systems theory benefit considerably from computer simulations (Polking, 1995; Scheinerman, 1995; Strogatz, 1994). Most relevant to the cognitive sciences, dynamical systems theory is being

successfully applied to a wide range of human behaviors, such as categorization (Anderson et al., 1977), language (Tabor & Tanenhaus, 1999), visual perception (Grossberg, 1980), motor movement (Kelso, 1995), as well as music perception (Large & Palmer, 2002), and developmental processes (Thelen & Smith, 1994). I genuinely suspect at this point that these advances of dynamical systems in various subfields of psychology spell doom for the computer metaphor of the mind.

As should be evident by now, the purpose of this book is to deprogram the cognitive psychologist in us all. We all have a tendency to want to draw a circle around a set of phenomena and label that set with a name like *perception* and perhaps label another set of phenomena with the name *cognition*. Even within those circles, we feel the need to draw smaller circles of things like "word recognition," as if it was completely unrelated to "object recognition." We all have a tendency to want to draw boxes around presumed transformations of information (e.g., combining spoken sounds over time to map onto words representations, or combining visual features and surfaces to map onto object representations), and call them *processors* or *modules*. We have these tendencies because without these overidealized categorical separations and discrete labels, we feel at a loss for how to talk about these phenomena. But how do I refer to a process that combines spoken sounds *and* visual features over time to map onto possible motor actions? The vocabulary of traditional cognitive psychology is simply not built for it. In contrast, the intersection of dynamical systems theory, neural network modeling, and ecological psychology, a nexus that I refer to as continuity psychology, is developing not only the vocabulary but also the conceptual and mathematical tools for it.

As we watch traditional cognitive psychology giving way to continuity psychology, one is tempted to ask, as Douglas Hintzman (1993) did, "Was the cognitive revolution a mistake?" And I think the answer is clearly "no"—but not because it got anything right about the mind. The cognitive revolution of the 1960s was the right thing to do at the time because, in opposing the antimentalism of the behaviorist tradition, it provided the necessary realization that the mind has sufficient complexity of processing to make it required reading, as it were. Psychology could no longer focus solely on the stimulus and the response, ignoring the complex nested dynamical processes that take place in between. The first-order associationism of the 1940s simply wasn't powerful enough to fit the data (Lashley, 1951; see also Chomsky, 1959). Unfortunately, where cognitive psychology in particular and cognitive science in general went wrong was in its marriage to the computer metaphor of the mind. Box-and-arrow diagrams, borrowed from computer engineering, ran amok in the scientific journals, and serial digital processes were used as the square pegs to be forced into the round holes of cognition. The mind was treated as an independent system, somehow composed of multiple internal independent subsystems.

However, in the past few decades, evidence from ecological psychology, neuroscience, and real-time methodologies in cognitive psychology has cast