

The Embodied Mind

Cognitive Science and Human Experience

revised edition

Francisco J. Varela

Evan Thompson

Eleanor Rosch

new foreword by Jon Kabat-Zinn

new introductions by Evan Thompson and Eleanor Rosch

The MIT Press
Cambridge, Massachusetts
London, England

PH/R
2017
5000(1, 2016)



© 1991, 2016 Massachusetts Institute of Technology

All rights reserved. No part of this book may be reproduced in any form by any electronic or mechanical means (including photocopying, recording, or information storage and retrieval) without permission in writing from the publisher.

This book was set in Stone Sans and Stone Serif by Toppan Best-set Premedia Limited.
Printed and bound in the United States of America.

Library of Congress Cataloging-in-Publication Data

Names: Varela, Francisco J., 1946–2001, author. | Thompson, Evan, author. | Rosch, Eleanor, author.

Title: The embodied mind : cognitive science and human experience / Francisco J. Varela, Evan Thompson, and Eleanor Rosch ; foreword by Jon Kabat-Zinn.

Description: Revised Edition. | Cambridge, MA : MIT Press, 2016. | Includes bibliographical references and index.

Identifiers: LCCN 2016016506 | ISBN 9780262529365 (pbk. : alk. paper)

Subjects: LCSH: Cognition. | Cognitive science. | Experiential learning. | Buddhist meditations.

Classification: LCC BF311 .V26 2016 | DDC 153–dc23 LC record available at <https://lccn.loc.gov/2016016506>

10 9 8 7 6 5 4 3 2 1

*Those who believe in substantiality are like cows;
those who believe in emptiness are worse.*

Saraha (ca. ninth century CE)

8 Enaction: Embodied Cognition

Recovering Common Sense

The tacit assumption behind the varieties of cognitive realism (cognitivism, emergence, and the society of mind) has been that the world can be divided into regions of discrete elements and tasks. Cognition consists in problem solving, which must, if it is to be successful, respect the elements, properties, and relations within these pre-given regions.

This approach to cognition as problem solving works to some degree for task domains in which it is relatively easy to specify all possible states. Consider for example the game of chess. It is relatively easy to define the constituents of the “space of chess”: there are positions on the board, rules for movements, turns that are taken, and so on. The limits of this space are clearly defined; in fact, it is an almost crystalline world. It is not surprising, then, that chess playing by computer is an advanced art.

For less circumscribed or well-defined task domains, however, this approach has proved to be considerably less productive. Consider, for example, a mobile robot that is supposed to drive a car within a city. One can still single out in this “driving space” discrete items, such as wheels and windows, red lights, and other cars. But unlike the world of chessplaying, movement among objects is not a space that can be said to end neatly at some point. Should the robot pay attention to pedestrians or not? Should it take weather conditions into account? Or the country in which the city is located and its unique driving customs? Such a list of questions could go on forever. The driving world does not end at some point; it has the structure of ever-receding levels of detail that blend into a nonspecific background. Indeed, successfully directed movement such as driving

depends upon acquired motor skills and the continuous use of common sense or background know-how.

Such commonsense knowledge is difficult, perhaps impossible, to package into explicit, propositional knowledge—"knowledge *that*" in the philosopher's jargon—since it is largely a matter of readiness to hand or "knowledge *how*" based on the accumulation of experience in a vast number of cases. Recent examinations of how skills are acquired appear to confirm this point.¹ Furthermore, when we enlarge the task domains from artificial microworlds to the world at large, it is not clear that we can even specify what is to count as an object independent of the type of action that is being performed. The individuation of objects, properties, and events appears to vary according to the task at hand.²

These points are not new to the field of cognitive science, although their full import has only begun to be appreciated. Indeed, it is fair to say that by the 1970s, after two decades of humblingly slow progress, it dawned on many workers in cognitive science that even the simplest cognitive action requires a seemingly infinite amount of knowledge, which we take for granted (it is so obvious as to be invisible) but which must be spoon-fed to the computer. The early cognitivist hope for a general problem solver had to be abandoned in favor of programs that would run in local knowledge domains, where small-scale problems could be solved and where the programmer could put into the machine as much of her background knowledge as was necessary. Similarly, the current connectionist strategy depends either on restricting the space of possible attractors by means of assumptions about the known properties of the world, which are incorporated as additional constraints for regularization,³ or, in more recent models, on using backpropagation methods where learning resembles the imitation of an external model. Thus in both cognitivism and connectionism, the unmanageable ambiguity of background common sense is left largely at the periphery of the inquiry, with the hope that it will somehow eventually be clarified.⁴

If, however, our lived world does not have predefined boundaries, then it seems unrealistic to expect to capture commonsense understanding in the form of a representation—where *representation* is understood in its strong sense as the re-presentation of a pre-given world.⁵ Indeed, if we wish to recover common sense, then we must invert the representationist attitude by treating context-dependent know-how not as a residual artifact

that can be progressively eliminated by the discovery of more sophisticated rules but as, in fact, the very essence of *creative cognition*.

This attitude toward common sense has begun to affect the field of cognitive science, especially in artificial intelligence. We should note, however, that the philosophical source for this attitude is to be found largely in recent Continental philosophy, especially in the school of philosophical hermeneutics, which is based in the early work of Martin Heidegger and his student Hans Gadamer.⁶ The term *hermeneutics* originally referred to the discipline of interpreting ancient texts, but it has been extended to denote the entire phenomenon of interpretation, understood as the *enactment* or *bringing forth* of meaning from a background of understanding. In general, Continental philosophers, even when they explicitly contest many of the assumptions underlying hermeneutics, have continued to produce detailed discussions that show how knowledge depends on being in a world that is inseparable from our bodies, our language, and our social history—in short, from our *embodiment*.⁷

Although several cognitive scientists have recently turned to these discussions for inspiration, the spontaneous philosophy of cognitive science continues to resist such a nonobjectivist orientation. The varieties of cognitive realism are in particular strongly tied to analytic philosophy, which tends to view folk psychology as a tacit theory in need of either reduction or replacement.⁸ Indeed, it is fair to say that analytic philosophy in general resists this notion of cognition as embodied understanding. Thus as Mark Johnson notes in a recent work,

The idea that understanding is an event in which one has a world, or, more properly, a series of ongoing related meaning events in which one's world stands forth, has long been recognized on the Continent, especially in the work of Heidegger and Gadamer. But Anglo-American analytic philosophy has steadfastly resisted this orientation in favor of meaning as a fixed relation between words and the world. It has been mistakenly assumed that only a viewpoint that transcends human embodiment, cultural embeddedness, imaginative understanding, and location within historically evolving traditions can guarantee the possibility of objectivity.⁹

The central insight of this nonobjectivist orientation is the view that knowledge is the result of an ongoing interpretation that emerges from our capacities of understanding. These capacities are rooted in the structures of our biological embodiment but are lived and experienced within a domain of consensual action and cultural history. They enable us to make sense of

our world; or in more phenomenological language, they are the structures by which we exist in the manner of “having a world.” To quote Johnson once more,

Meaning includes patterns of embodied experience and preconceptual structures of our sensibility (i.e., our mode of perception, or orienting ourselves, and of interacting with other objects, events, or persons). These embodied patterns do not remain private or peculiar to the person who experiences them. Our community helps us interpret and codify many of our felt patterns. They become shared cultural modes of experience and help to determine the nature of our meaningful, coherent understanding of our “world.”¹⁰

Although these themes are derived from Continental philosophy, most of the Continental discussions have proceeded without taking into consideration scientific research on cognition—the major exception being the early work of Merleau-Ponty. The challenge posed by cognitive science to the Continental discussions, then, is to link the study of human experience as culturally embodied with the study of human cognition in neuroscience, linguistics, and cognitive psychology. In contrast, the challenge posed to cognitive science is to question one of the more entrenched assumptions of our scientific heritage—that the world is independent of the knower. If we are forced to admit that cognition cannot be properly understood without common sense, and that common sense is none other than our bodily and social history, then the inevitable conclusion is that knower and known, mind and world, stand in relation to each other through mutual specification or dependent coorigination.

If this critique is valid, then scientific progress in understanding cognition will not be forthcoming unless we start from a different basis from the idea of a pre-given world that exists “out there” and is internally recovered in a representation. In recent years, a few researchers within cognitive science have taken this critique from the philosophical level into the laboratory and into specific work in AI. These researchers have put forth concrete proposals that involve a more radical departure from cognitivism than is found in the emergence approach, and yet they incorporate the ideas and methods developed within this context.

Self-Organization Revisited

In the previous chapter, we discussed how cognitive science has slowly drifted away from the idea of mind as an input-output device that

processes information toward the idea of mind as an emergent and autonomous network. We intend to make this idea more tangible by providing a concrete example of what we mean by an autonomous system.

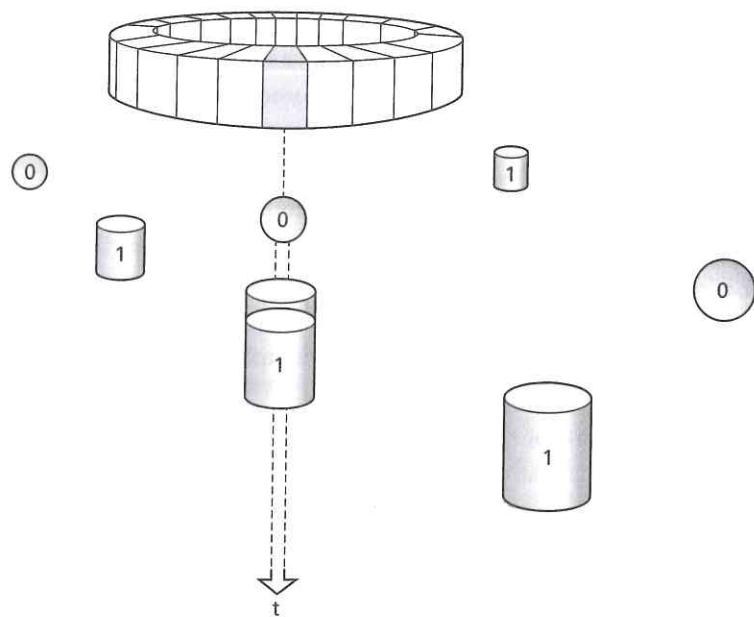
Our example is based on the simple cellular automata, which we introduced to exemplify how systems exhibit emergent properties when endowed with network architectures. In the previous account these cellular automata were completely decoupled entities, and so their emergent states were not constrained by a history of coupling with an appropriate world. By enriching our account to include this dimension of *structural coupling*, we can begin to appreciate the capacity of a complex system to enact a world.¹¹

There are many forms of coupling that we could provide for our rings. Let us suppose, though, that we simply drop the ring into a milieu of random 0s and 1s, much like a cell that is plunged into a chemical milieu. Imagine further that when one of the cells of this automata encounters one of these two alternatives (0s and 1s), the state of the cell is replaced by the perturbation that it encountered (see figure 8.1). For the sake of brevity, let us give the name *Bittorio* to this particular ring of cellular automata having this form of structural coupling with the chosen milieu.

In figure 8.2, the arrow to the left indicates the moment where one perturbation reaches one particular cell at one particular instant. The dynamics that follow indicate the ensuing change (or lack of it), that is, the way in which Bittorio compensates for this perturbation. If Bittorio’s rule belongs to the first or fourth class (a simple or a chaotic attractor), then the consequence of the perturbation is simply invisible: Bittorio either goes back to its previous homogenous state, or it remains in a randomlike state.

It follows that only the second and third classes of rules can provide us with dynamics capable of producing interesting consequences for the kind of structural coupling we have chosen for Bittorio. As figure 8.2 shows, for Bittorios with these rules a single perturbation induces a change from one to another spatiotemporal configuration. Both of these configurations are stable and distinguishable.

The case of Bittorio of rule 10010000, illustrated in figure 8.3, is worth commenting on in more detail. As can be seen, the encounter with just one perturbation changes the spatial periodicities from one to another stable configuration. But a second perturbation at the same cell undoes the previous change. Hence any odd sequence of perturbations at the same locus

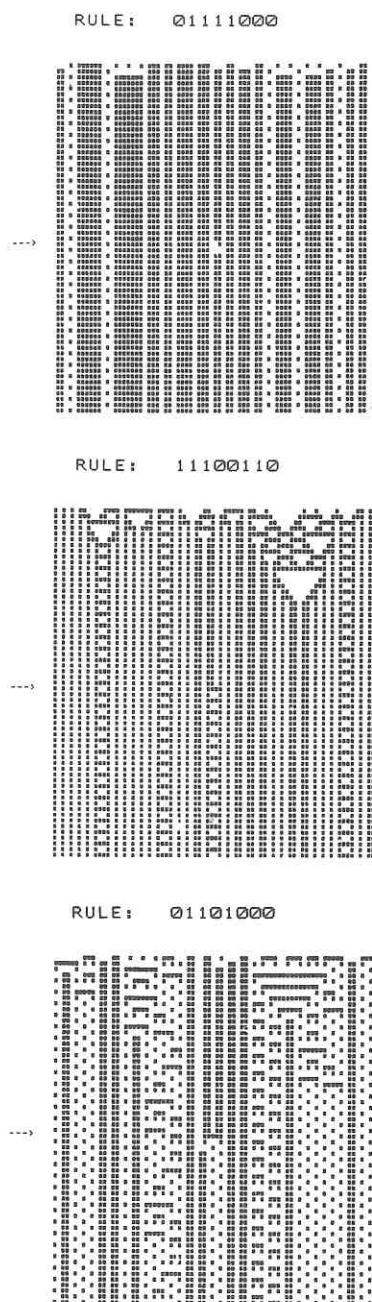
**Figure 8.1**

Cellular automata Bittorio in a random soup of 1s and 0s.

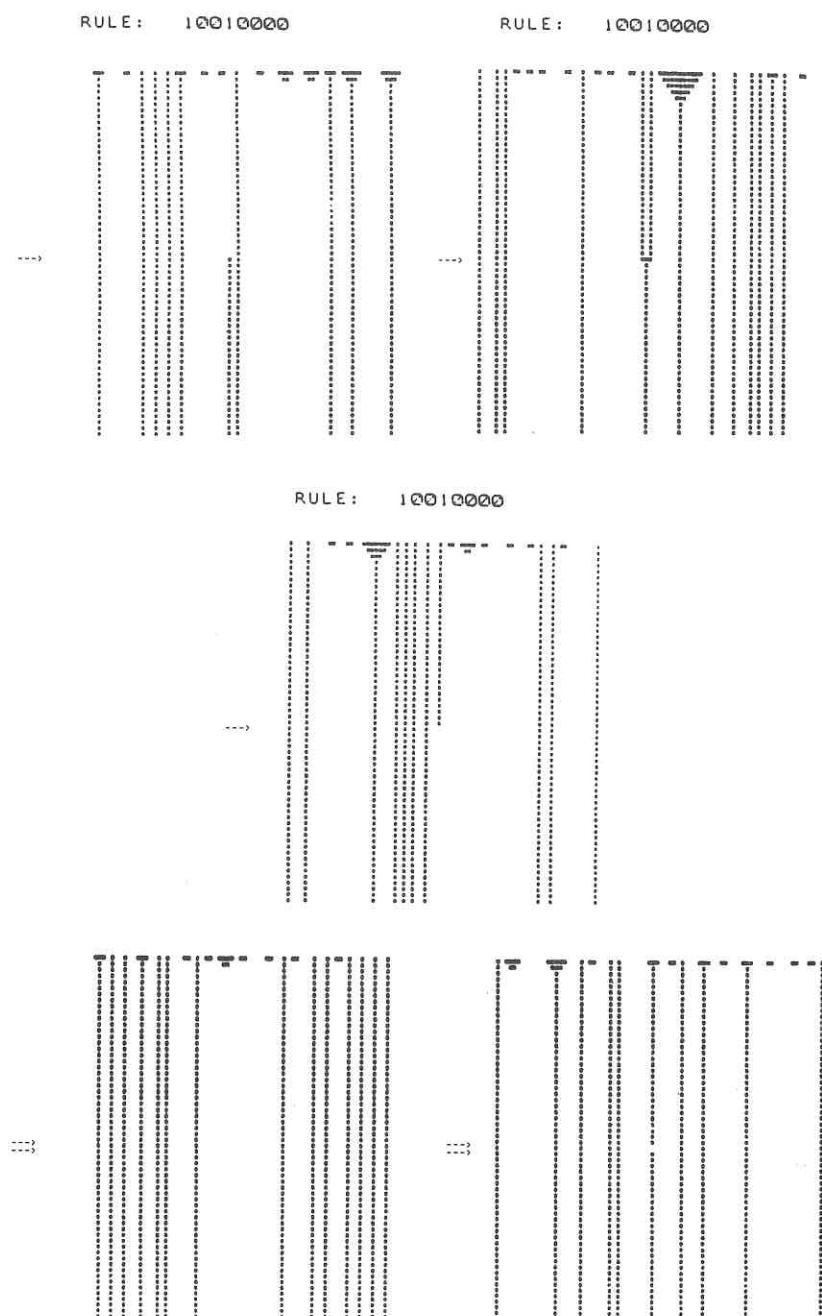
will lead to a change in state configuration for Bittorio, whereas any even sequence of perturbations will be invisible since it leaves Bittorio unchanged. Thus of all the innumerable sequences of possible perturbations, this Bittorio picks up or singles out from the milieu a very specific subset, namely, finite odd sequences, since only these sequences induce a repeatable change in Bittorio's configuration. In other words, given its rule and given its form of structural coupling, this Bittorio becomes an "odd sequence recognizer."

Another example of such emergent significations is shown in figure 8.4 for Bittorio of rule 01101110. Here a sequence of two perturbations is the only trigger capable of leading to a change in the state configuration of Bittorio. This is readily seen in figure 8.4 where we have superimposed several encounters at different cellular loci to facilitate comparison. Anything other than double perturbations in one location leaves this Bittorio unchanged.

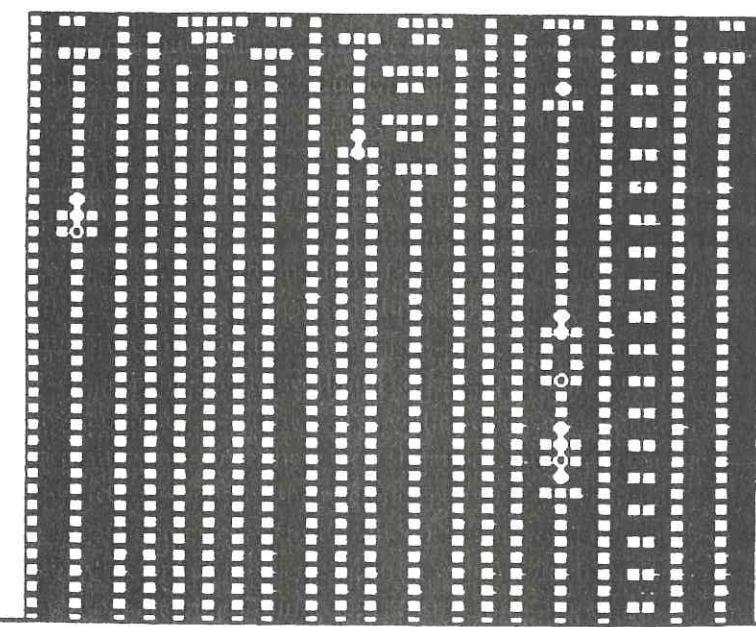
Other explorations with simultaneous perturbations and more complex forms of structural coupling reveal more rich and interesting behaviors for

**Figure 8.2**

Bittorio's life history showing changes in this history depending on the perturbations it encounters.

**Figure 8.3**

A Bittorio of rule 10010000, choosing only odd sequences of perturbations.

**Figure 8.4**

A Bittorio responsive to a sequence of double perturbations.

these Boolean cellular automata. The above examples, however, are enough for the purposes of illustration here.

We wish to emphasize that in these two specific cases (figures 8.3 and 8.4) we have not provided Bittorio with a program to distinguish "odd sequences" or "two successive perturbations." Instead we have specified, on the one hand, a form of closure for the system (the network's internal dynamical emergences) and, on the other hand, the way in which this system will couple with a given milieu (replacement of the state of each cell with the perturbation it encounters in a milieu of random 0s and 1s). The result, however, is that over time this coupling selects or enacts from a world of randomness a domain of distinctions ("odd sequences" or "two successive perturbations") that has relevance for the structure of the system. In other words, on the basis of its autonomy the system selects or enacts a domain of significance.

We use these words *significance* and *relevance* advisedly, for they imply that there is some kind of interpretation involved in the encounters. In the

case of Bittorio, this interpretation is obviously a far cry from the kinds of interpretation that depend on experience. Nevertheless, we can say that a minimal kind of interpretation is involved, where *interpretation* is understood widely to mean the enactment of a domain of distinctions out of a background. Thus Bittorio, on the basis of its autonomy (closure), performs an interpretation in the sense that it selects or brings forth a domain of significance out of the background of its random milieu.

The distinctions that Bittorio selects, such as odd sequences, indicate the regularities with which Bittorio covaries. These regularities constitute what we could call Bittorio's world. It should be apparent that this world is not pre-given and then recovered through a representation. We did not design Bittorio to be an odd sequence recognizer; we simply provided Bittorio with certain internal dynamics and then dropped it into a random milieu. Nevertheless, given the history of coupling between the internal dynamics and the milieu, *odd sequence* becomes a significant distinction for Bittorio. For this reason, we describe Bittorio's world as enacted through a history of structural coupling.

Bittorio provides, then, a paradigm for how closure and coupling suffice to bring forth a world of relevance for a system. Of course, this paradigm is rather simple. Our intention, however, is not to provide a model of any specific phenomenon, and we certainly do not intend to suggest that such a simple form of closure and coupling is sufficient for a system to experience a world. Rather, our intention is simply to provide a *minimal* example of how an autonomous system brings forth significance from a background. It is the simplicity of the example that enables us to follow in detail the entire process by which a kind of distinction is enacted.

Despite the simplicity of the example, we should not underestimate the moral it suggests. Since we can already recognize the emergence of a minimal kind of significance with just the simple form of autonomy (closure) and coupling given to Bittorio, imagine the rich and complex kinds of significance that would be brought forth by living cells or complex cellular networks, such as the brain and the immune system. Though far more complex and intricate, these systems nonetheless share with Bittorio the properties of being autonomous (having operational closure) and being structurally coupled.¹²

Such autonomous systems stand in sharp contrast to systems whose coupling with the environment is specified through input/output relations.

The digital computer is the most familiar example of this latter kind of system. Here the meaning of a given keyboard sequence is always assigned by the designer. Living systems, however, are far from being in this category. Under very restricted circumstances we can speak as if we could specify the operation of a cell or an organism through input/output relations. In general, though, the meaning of this or that interaction for a living system is not prescribed from outside but is the result of the organization and history of the system itself. Let us now turn, then, to consider some actual living examples.

Color as a Study Case

Perhaps the best example, one which we intend to explore in some depth here, is color perception. We have two reasons for choosing to focus on color. First, the study of color provides a microcosm of cognitive science, for each discipline in figure 1.1—neuroscience, psychology, artificial intelligence, linguistics, and philosophy—has made important contributions to our understanding of color. Indeed, other disciplines, such as genetics and anthropology, have contributed as well. Second, color has immediate perceptual and cognitive significance in human experience. For these two reasons, color provides a paradigmatic domain in which our twin concerns of science and human experience naturally intersect.

For ease of exposition our discussion of color will proceed through several stages. We will first discuss how colors themselves appear—what could be called the structure of color appearance. We will then discuss color as a perceived attribute of things in the world. Finally, we will consider color as an experiential category. Let us emphasize that these stages are not found separately in experience; our experience is simultaneously shaped by all three. Theories of color do, however, tend to take as their point of departure one or the other of these three aspects. Thus our stages, though expository, are not arbitrary.

Color Appearance

Let us begin, then, not with the visual system or with colored objects but simply with color itself. There are two important features of the structure of color appearance. First, all of the colors that we see can be described as some combination of six basic colors: red, green, yellow, blue, black, and

white. For example, orange is a combination of red and yellow; turquoise is a combination of blue and green; violet and indigo are combinations of red and blue, etc. Second, the appearance of color varies along three dimensions, those of hue, saturation, and brightness. *Hue* refers to the degree of redness, greenness, yellowness, or blueness of a given color. Red, green, yellow, and blue are the four fundamental or psychologically unique hues, which combine to form complex or psychologically binary hues. For example, red and yellow combine to form reddish-yellows and yellowish-reds (oranges), whereas blue and red combine to form blueish-reds and reddish-blues (purples). For each unique hue, there is another unique hue with which it cannot coexist to form a binary hue. Thus red cannot coexist with green, and yellow cannot coexist with blue. Red and green are therefore known as *opponent* hues, as are blue and yellow. It should be noted that not every color need be of a certain hue. White and black, as well as the intermediate shades of gray, are colors, but they have no hue. They are therefore known as achromatic colors—colors that have zero hue—whereas colors with hue are called chromatic. The chromatic colors can also differ in the strength or *saturation* of their hue. Saturated colors have a greater degree of hue, whereas desaturated colors are closer to gray. *Brightness* is the final dimension of color appearance. Along this dimension, colors vary from dazzling at one end to dim or barely visible at the other end.

Why does color have this structure? Why, for example, are hues organized into pairs that are mutually exclusive or opponent? The model of color vision that takes as its point of departure the structure of color appearance, and so attempts to answer these questions, is known as the *opponent-process theory*. This theory owes its origin to the research of the nineteenth-century physiologist Ewald Hering but was proposed in its modern form by Leo Hurvich and Dorothea Jameson in 1957.¹³ According to this theory, there are three color “channels” in the visual system: one channel is achromatic and signals differences in brightness; the other two are chromatic and signal differences in hue. It should be noted that these channels are specified in psychophysical experiments, not neurophysiological ones. The exact nature of their physiological embodiment is still a matter of debate. Nevertheless, it is accepted that the channels correspond in some way to the complex cross-connections among retinal cells and postretinal neuronal ensembles.

In the retina there are three different but intermingled mosaics of cone cells, whose overlapping photopigment absorption curves peak around 560, 530, and 440 nanometers respectively. These three cone mosaics constitute the so-called long-wave (L), middle-wave (M), and short-wave (S) receptors. Excitatory and inhibitory processes in postreceptoral cells enable the signals from these receptors to be added and/or subtractively compared. In the opponent-process model, the addition of the signals from all three receptors generates the achromatic (brightness) channel. The difference between the signals from the L and M receptors generates the red-green channel, and the difference between the sum of the signals from the L and M receptors and the signals from the S receptors generates the yellow-blue channel. These two chromatic channels are opponent: an increase in red is always gained at the expense of green and vice-versa; an increase in yellow is always gained at the expense of blue and vice-versa.

This opponent-process theory explains the structure of color appearance by showing how it results from the differential responses of the achromatic and chromatic channels. Thus the organization of hues into mutually exclusive or antagonistic pairs reflects an underlying opponent organization. We never experience any color to be a combination of red and green, or yellow and blue, because the chromatic channels cannot simultaneously signal “red” and “green,” or “yellow” and “blue.” The opponent-process theory also explains why some hues are unique and others are binary. Unique hues result from a signal from one chromatic channel while the other chromatic channel is neutral or balanced. For example, unique green results when the red-green channel signals “green” and the yellow-blue channel is neutral so that it signals neither “yellow” nor “blue.” Binary hues, on the other hand, result from the interplay of the two channels with each other. Thus orange results from the red-green channel signaling “red” and the yellow-blue channel signaling “yellow.”

Now that we have a basic understanding of how color appearance is generated, let us turn to the second stage in our investigation, color as a perceived attribute of things in the world.

Color as a Perceived Attribute

Since we perceive colors to be spatially located, we might assume that the color we perceive an area to have can be correlated with the light reflected locally from that area. Thus if some area looks whiter than another, it must

be because more light is reflected from the area. Or if some area looks green, it must be that the area reflects predominantly middle-wave light. If we fail to see the area as green in such a situation, then our perception must be mistaken; what we see must be an illusion.

If we examine the situation more closely, however, we are in for interesting surprises. If we actually measure the light reflected from the world around us, we will discover that there simply is no one-to-one relationship between light flux at various wavelengths and the colors we perceive areas to have. Suppose, for example, that we perceive some area to be green. Areas that look green typically reflect a high percentage of middle-wave light and a low percentage of long-wave and short-wave light. We might suppose, then, that the area looks green because it reflects more middle-wave light to the eye. This supposition would be true, however, only in the limited case where the area is viewed in isolation, that is, if we exclude everything else from the field of view. But when this area is viewed as part of a complex scene, it will continue to look green even if it reflects more long-wave and short-wave light than middle-wave light. In other words, when the area is viewed as part of a complex scene, the light that it locally reflects is not sufficient to predict its perceived color. Therefore, there simply is no one-to-one correspondence between perceived color and locally reflected light.

This relative independence of perceived color from locally reflected light has been known to vision scientists for quite some time.¹⁴ The independence is manifested in two complementary phenomena. In the first, the perceived colors of things remain relatively constant despite large changes in the illumination. This phenomenon is known as *approximate color constancy*. In the second, two areas that reflect light of the same spectral composition can be seen to have different colors depending on the surroundings in which they are placed. This phenomenon is known as *simultaneous color contrast* or *chromatic induction*.¹⁵

These two phenomena force us to conclude that we cannot account for our experience of color as an attribute of things in the world by appealing simply to the intensity and wavelength composition of the light reflected from an area. Instead, we need to consider the complex and only partially understood processes of cooperative comparison among multiple neuronal ensembles in the brain, which assign colors to objects according to the emergent, global states they reach given a retinal image.

Consider the following interesting demonstration. We take two identical slide projectors superimposed over a common screen and fit each with identical copies of a slide containing a checkerboard of grays, whites, and blacks. The two slides are superimposed so that they are exactly aligned. We also put a red filter in one of the projectors, so that the overall pattern that results is an array of pinks of different saturations. Let us now turn one slide by 90 degrees. The result is a full, multicolored image, containing small squares that are yellow, blue, and green, as well as red and pink.¹⁶

The effect of this experiment is quite dramatic: a multicolored image arises where physics would lead us to expect only various shades of pink. This chromatic effect can be described by the white-to-white and red-to-red ratios across the edges of the small squares accomplished by the rotation of one of the slides. How can this happen?

As we mentioned when discussing the opponent-process theory, the light that reaches the eye perturbs three different but intermingled mosaics of cones, which constitute three retinal surfaces: the S, M, and L receptors. These three retinal surfaces are by no means identical or homogenous. For example, the L receptor has a density of cones about five times higher than the S receptor and slightly less than the M receptor. Furthermore, due to the inner connectivity of the retina, local differences of activity in the three receptor surfaces depend on what happens in the rest of the retina. In this manner, internal *relative* values are generated. Abrupt deviations from such reference values in the local levels of activity become the difference that makes a difference: within the boundaries of such deviations a uniform color is perceived.

This description highlights the emergent configurations at the retinal level and so is only partial. There are structures at all levels of the visual pathways that participate in the perception of color. In primates, the participation of subensembles of neurons in color perception has been demonstrated in the thalamus (LGN), primary and extrastriate visual cortex, inferotemporal cortex, and frontal lobes.¹⁷ Most notable is a collection of neurons in the so-called area V4 of the extrastriate cortex where even individual neuronal responses can be roughly associated with the color constancies of a visual field.¹⁸ These neuronal structures constitute a color subnetwork—a sort of perceptual “agent,” to use Minsky’s terminology. Thus nothing short of a large and distributed neuronal network is involved in our perception of color.

Colors are not, of course, perceived in isolation from other attributes, such as shape, size, texture, motion, and orientation. For example, the artist Kandinsky commented on the relation between color and motion. In one of his essays he wrote, "If two circles are drawn and painted respectively yellow and blue, a brief contemplation will reveal in the yellow a spreading movement out from the center, and a noticeable approach to the spectator. The blue, on the other hand, moves into itself, like a snail retreating into its shell, and draws away from the spectator. The eye feels stung by the first circle while it is absorbed into the second."¹⁹

The motion that Kandinsky refers to here is obviously not movement in the physical space of the picture. It is, rather, motion in our perceptual space. As Mark Johnson notes in a discussion of this passage from Kandinsky, "The 'movement' refers to structures in our perceptual interaction, in which we form unified images and trace out relations among the various elements in the work."²⁰

Recent trends in physiology enable us to understand the bodily basis for these "structures of perceptual interaction." In recent years physiology has moved toward the study of vision as a patchwork of visual modalities, including at least form (shape, size, rigidity), surface properties (color, texture, specular reflectance, transparency), three-dimensional spatial relationships (relative positions, three-dimensional orientation in space, distance), and three-dimensional movement (trajectory, rotation). It has become evident that these different visual modalities are emergent properties of concurrent subnetworks, which have a degree of independence and even anatomical separability but cross-correlate and work together so that at almost every moment a visual percept is coherent.²¹ (This kind of architecture is, once again, strongly reminiscent of Minsky's societies of agents.) Figure 8.5 depicts some of the identified anatomical elements of these visual subnetworks. Among the modalities, color seems to be one of the more simple, for color indicators can be obtained solely on the basis of luminance and contrast levels. This simplicity betrays, however, the equally important fact that color is always perceived within a more encompassing visual context. All of the subnetworks work cooperatively; we never see color as an isolated item.

Furthermore, visual perception is in active exchange with other sensory modalities. For example, the associations of color and sound, as well as color and horizontal/vertical perception (involving the sense of orientation

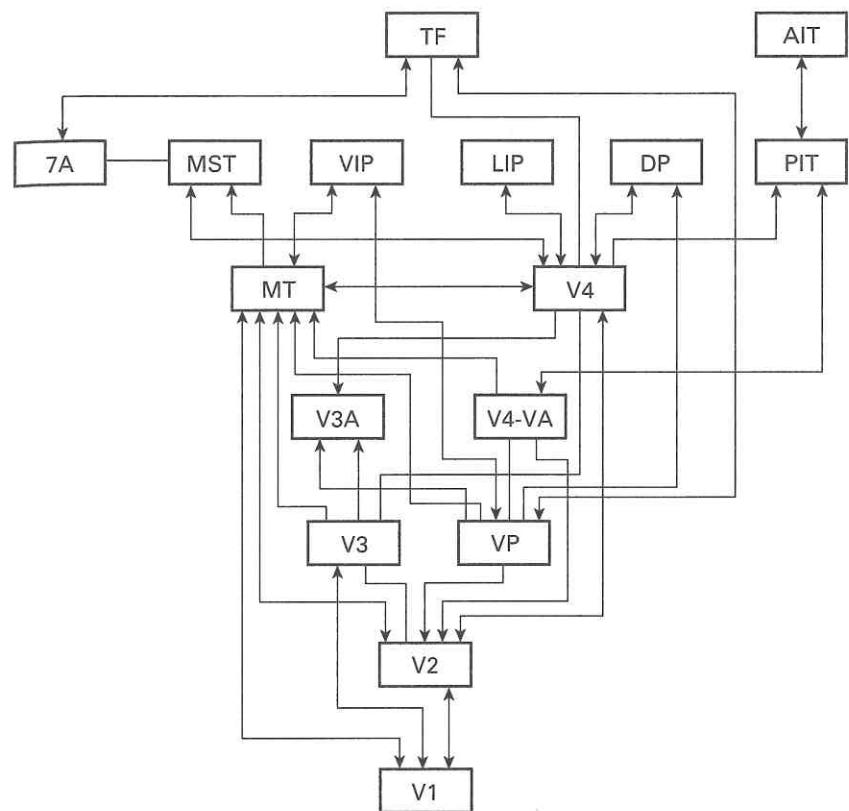


Figure 8.5

Parallel streams in the visual pathway. From DeYoe and Van Essen, Concurrent processing streams in monkey visual cortex.

and equilibrium), are well known to artists, though less studied by neurobiologists. Beyond these intermodal relations there are, of course, varieties of cognitive expectancies and memories. Such "top-down" dependency is to be expected, for, like those of the LGN and visual cortex, the pathways depicted in figure 8.5 are all bidirectional. Thus, to reiterate one of our central points, the neuronal network does not function as a one-way street from perception to action. Perception and action, sensorium and motorium, are linked together as successively emergent and mutually selecting patterns.

To bring home this point that color perception partakes of both other visual and sensory modalities, let us consider a much more dramatic

example: the complete loss of color perception. In a recent article Oliver Sacks and Robert Wasserman presented an account of a patient who, due to an accident, became completely color-blind.²² This particular case of so-called acquired cerebral achromatopsia is fascinating because it occurred in an artist known for his especially colorful, abstract paintings. As a result of a car accident, this person—referred to as “Mr. I”—could no longer perceive any colors: he lived in a visual world that resembled black-and-white television.

The participation of color perception in other modalities of experience is evident from Mr. I’s descriptions in the weeks following his accident. Because of the absence of color, the overall character of his experience changed dramatically: everything he saw “had a distasteful, ‘dirty’ look, the whites glaring, yet discolored and off-white, the blacks cavernous—everything wrong, unnatural, stained, and impure.”²³ As a result, he found foods disgusting and sexual intercourse impossible. He could no longer visually imagine colors, nor could he dream in color. His appreciation of music was also impaired, for he could no longer experience musical tones by synesthetically transforming them into plays of color. Eventually, Mr. I seemed to forget completely his former world of color. His habits, behavior, and actions changed as he became progressively more of a “night person.” In his words, “I love the nighttime. ... I often wonder about people who work at night. They never see the sunlight. They prefer it. ... It’s a different world: there’s a lot of space—you’re not hemmed in by streets, people. ... It’s a whole new world. Gradually I am becoming a night person. At one time I felt kindly toward color, very happy about it. In the beginning, I felt very bad, losing it. Now I don’t even know it exists—it’s not even a phantom.”²⁴

This description provides rare insight into how our perceived world, which we usually take for granted, is constituted through complex and delicate patterns of sensorimotor activity. Our colored world is brought forth by complex processes of structural coupling. When these processes are altered, some forms of behavior are no longer possible. One’s behavior changes as one learns to cope with new conditions and situations. And, as one’s actions change, so too does one’s sense of the world. If these changes are dramatic enough—as in Mr. I’s loss of color—then a different perceived world will be enacted.

The preceding examples have shown us how color as an attribute is intimately involved with other attributes of our perceived world. Our examination so far shows that we will not be able to explain color if we seek to locate it in a world independent of our perceptual capacities. Instead, we must locate color in the perceived or experiential world that is brought forth from our history of structural coupling. Indeed, this point will become even more apparent when we consider color as an experiential category. Before we turn to this third stage in our discussion of color, however, let us pause to consider an objection.

Where Is Color?

Suppose someone, in reply to our discussion, were to demand, “What is the point of all these complex neuronal processes if not to compensate for the changes in illumination and recover some stable feature of objects? Consider, for example, the surface reflectance of an object. This property corresponds to the percentage of incident light at each wavelength that an object reflects. This percentage or ratio describes the way in which an object, by virtue of its physical constitution, alters the ambient light; it is therefore a stable property, one that remains constant through changes in illumination. Why not say, then, that although we must account for color experience by revealing its constitution through emergent patterns of neuronal activity, this experience is nonetheless a result of having to solve the information-processing problem of recovering surface reflectance?”

Recent computational models of color vision seem to support this line of argument. The surface reflectances of objects in our surrounding world, such as bricks, grass, buildings, etc., can be expressed in a rather limited (three-dimensional) set of prototypical functions.²⁵ Thus it would seem that all the visual system has to do is sample the scene with its three color channels and thereby reconstitute the surface reflectances from the activity in these channels. On the basis of these models, several vision scientists, as well as certain philosophers, have argued not only that the function of color vision is the recovery of surface reflectance but also that color itself is just the property of surface reflectance.²⁶

This objectivist proposal gives rise to several considerable problems, which serve to reinforce our point that the colors we see must be located not in a pre-given world but rather in the perceived world brought forth from our structural coupling. Consider first the idea that color is just

surface reflectance. We have already seen that colors have certain properties and bear certain relations to each other: color varies along the three dimensions of hue, saturation, and brightness; hues are either unique or binary and are organized into opponent pairs, etc. Now if color is just surface reflectance, we should be able to match these features of color with corresponding features of surface reflectance. But there are no such corresponding features. Surface reflectances can be classified according to whether they reflect more or less light in the short-, middle-, and/or long-wave regions of the spectrum, but they cannot be classified as being unique or binary, nor can they be classified as standing in opponent relations to other reflectances. Nor can these properties of uniqueness, binariness, and opponency be found in the structure of light. For these reasons, the properties that specify what colors are simply have no nonexperiential, physical counterparts.²⁷

Second, color is not simply a perceived attribute of surfaces; it is also a perceived attribute of volumes such as the sky. Furthermore, we experience colors as attributes of afterimages and in dreams, memories, and synesthesia. The unity among these phenomena is not to be found in some nonexperiential, physical structure but rather in color as a form of experience that is constituted through emergent patterns of neuronal activity.

Let us now consider the idea that the function of color vision is to represent and thereby recover surface reflectance. The first thing to note about this idea is that it arises not from the biological and ecological investigation of color vision but from the engineering attempt to devise a system that will be able to detect objects by discounting variations in the illumination and recovering the invariant reflectances in a scene. Although this engineering research program is of considerable importance for our understanding of the more abstract principles involved in vision, it should not be allowed to dictate conclusions about the biological and ecological purposes that natural color vision serves. Indeed, attention to these biological and ecological purposes reveals that color vision is concerned as much with properties that change, such as lighting, weather conditions, and time of day, as with properties that remain constant, such as surface reflectance.²⁸

Finally, there is a hidden, but much deeper problem with the objectivist view of color vision: the objectivist simply assumes that surface reflectances are to be found in some pregiven world that is independent of our

perceptual and cognitive capacities. But how are we to specify what counts as a surface? How are we to specify its edges, boundaries, texture, and orientation, if not in relation to some perceiver for whom these distinctions are relevant?

The objectivist supposition that surface reflectances are pregiven rests on the assumption that since surface reflectance is a physical property, it can be measured and specified in entirely physical terms. But although the reflectance at any point in a scene can be specified in physical terms, what counts as a surface may in fact involve tacit reference to a type of perceiver. This point is obscured in computational models that emphasize the limited dimensions in which so-called naturally occurring reflectances can vary. If we actually examine these models, we will see that the natural reflectances correspond not only to the reflectances of typical objects from our human environment, as opposed to the environments of considerably different visual creatures, but also that these objects have been picked out or specified prior to the actual task of vision. In other words, these models treat the visual system as if it were simply presented with a certain class of prespecified objects whose reflectances must then be recovered.

This approach involves a considerable and artificial simplification of our actual perceptual situation. The visual system is never simply presented with pregiven objects. On the contrary, the determination of what and where an object is, as well as its surface boundaries, texture, and relative orientation (and hence the overall context of color as a perceived attribute), is a complex process that the visual system must continually achieve. This achievement, as we have seen in our discussion of the patchwork architecture of vision, results from a complex cooperative process involving active dialogue among all the visual modalities. Indeed, color vision is actually involved in the cooperative processes by which the visual scene comes to be segmented into a collection of surfaces. In the words of P. Gouras and E. Zrenner, "It is impossible to separate the object sensed from its color because it is the color contrast itself that forms the object."²⁹ Thus colors and surfaces go together: both depend on our embodied perceptual capacities.

Color as a Category

Our discussion so far has concentrated on the perception of color, considered either on its own terms (color appearance), as it were, or as an attribute

of things (surface colors, volume colors, etc.). But our experience of color is not only perceptual; it is also cognitive: we organize all the various hue/saturation/brightness combinations that we perceive into a limited set of color categories and give names to these categories. As we will now see, color categories provide yet another dramatic illustration of how color is brought forth.

Linguistic Aspects of Color Consider the numerous names that we have in English for colors: red, yellow, orange, green, blue, purple, violet, indigo, pink, turquoise, aquamarine, mauve, chartreuse, etc. Given these many names, as well as the numerous names in other languages, we might suppose that color categories are ultimately arbitrary, that is, that nothing compels us to categorize colors in one way rather than another. Indeed, this view was at one time dominant within the fields of linguistics and anthropology.³⁰

This view was dramatically challenged in 1969 with the publication of a now classic work by Brent Berlin and Paul Kay.³¹ In this work, Berlin and Kay specified a set of linguistic criteria for determining which color names in a given language constitute “basic” color terms. These basic color terms name the basic color categories in a given language. Then, in an examination of over ninety languages, Berlin and Kay determined that there are at most eleven basic color categories encoded in any language, though not all languages encode all eleven. These basic categories are red, green, blue, yellow, black, white, gray, orange, purple, brown, and pink. Berlin and Kay also presented speakers of various languages with a standardized array of color chips and asked them to specify both the boundaries and the best examples of the colors to which their basic terms refer. They found that although there was considerable variation among speakers over color category boundaries, individuals virtually always agreed on the best example of a color category. Furthermore, they found that when several languages contained a common basic term, such as a basic term for blue, speakers virtually always agreed on the best example of the color category no matter which language they spoke. Berlin and Kay argued therefore that the basic color categories do not have a uniform structure, for some members of the categories are central and so constitute category “foci.” Since these central members are universally agreed upon, Berlin and Kay concluded that “the eleven basic color categories are pan-human perceptual universals.”³²

Although some languages do not encode all eleven basic color categories, we should not suppose that the color domain is impoverished for speakers of these languages. On the contrary, the set of basic color terms in a given language always encompasses the entire color space. For example, the language of the Dani tribe of New Guinea has only two basic color terms. In studies of the Dani, Rosch (then Heider) showed that these two terms, which had previously been translated as “white” and “black,” were actually better translated as “white-warm” and “dark-cool,” for the former term covered white plus all the warm colors (red, yellow, orange, reddish-purple, pink), whereas the latter covered black plus all the cool colors (blue, green).³³

Color and Cognition The studies we have discussed so far have been about color language. There is an entire subfield in psychology, called language and cognition, that considers and disputes the ways in which language and cognition may or may not be related. Prior to Berlin and Kay, a well-known series of experiments had demonstrated that memory for colors (a cognitive variable) was a function of color naming (a linguistic variable).³⁴ Since naming was assumed to be culturally relative, it was thus argued and widely accepted that cognition had been demonstrated to be culturally relative. But what if both color language and color cognition are functions of some third underlying factor—color physiology, for example? A natural laboratory for testing such questions was provided by the Dani of New Guinea since their language lacked virtually all color vocabulary. In a series of experiments, Rosch found that (1) central members of basic color categories were perceptually more salient, could be learned more rapidly, and were more easily remembered in both short-term and long-term memory than were peripheral colors, even by speakers of Dani who do not have names for the central colors; (2) the structures of the color spaces derived from Dani and English color naming were very different but were quite similar for those derived from Dani and English color memory; and (3) when Dani were taught basic color categories, they found it quite easy to learn categories that were structured in the universal fashion (with central members as central) but extremely difficult to learn categories that were structured in a deviant manner (with peripheral colors as central, where blue-green might be central and blues and greens peripheral).³⁵ Very similar effects were found in the development of color names in young children in our

own culture.³⁶ All of these results argued strongly that both cognitive and linguistic aspects of color categorization are related to underlying (probably physiological) factors. Thus color categories appear to be a panhuman, species-specific universal.

Our discussion so far would seem to suggest that color categories are entirely determined by emergent patterns of neuronal activity in the human visual system—the color subnetwork that we reviewed above. Thus notice that the focal colors red, green, blue, yellow, black, and white can be mapped directly onto the responses of the three color channels in the opponent-process theory of color vision. But what about focal orange, purple, brown, and pink? More recent research suggests that distinctly cognitive operations are required to generate these focal colors. The cognitive operations appear to be of two kinds: one is universal for our species and the other is culture specific.³⁷

In 1978 Paul Kay and Chad McDaniel provided a model of how color categories could be generated from a certain set of neuronal responses plus certain species-specific cognitive processes.³⁸ The neuronal responses correspond to the red-green, yellow-blue, and black-white responses of neuronal ensembles, such as those found by R. DeValois and G. Jacobs in the LGN of the macaque, a species of monkey that has color vision quite similar to ours.³⁹ (One could also construct a model using the psychophysical color channels. Indeed, it is perhaps preferable to do so, since the exact neural embodiment of these channels is still disputed.) The cognitive processes correspond to operations that can be modeled by using a branch of mathematics known as fuzzy set theory. Unlike standard set theory, fuzzy set theory operates with sets that admit degrees of membership. Degree of membership in a set is specified by a function that assigns to each member of the set some value between 0 and 1. Thus for color, focal colors have degree of membership 1 in their respective categories, whereas nonfocal colors have degrees of membership between 0 and 1. In Kay and McDaniel's model, the red-green, yellow-blue, and black-white neuronal responses directly determine the basic categories red, green, yellow, blue, black, and white. Orange, purple, brown, and pink, however, are "computed" or "generated" by cognitive operations on these neuronal responses. These cognitive operations correspond to the operation of fuzzy set intersection. Thus orange is the fuzzy intersection of red and yellow, purple of red and blue, pink of white and red, and brown of black and yellow. Since these

categories require such cognitive derivations, Kay and McDaniel term them *derived* basic color categories.

Color and Culture Finally, color categories depend on culture-specific cognitive processes. Thus in another study, Paul Kay and Willett Kempton found that the lexical classification of colors can affect subjective judgments of similarity among colors.⁴⁰ For example, English contains terms for both green and blue, whereas Tarahumara (a Uto-Aztecan language of northern Mexico) has a single term that means "green or blue." This linguistic difference appears to be correlated with a difference in subjective judgments of similarity among colors between speakers of the two languages: English speakers tend to exaggerate the perceived distances of colors close to the green-blue boundary, whereas speakers of Tarahumara do not.

Other evidence for culture-specific cognitive processes comes from R. E. MacLaury. He has found that purple is sometimes placed entirely within the cool range (blue-green) and other times on the boundary between the cool range and red and that brown is sometimes placed within the yellow category and other times within black.⁴¹ MacLaury also reports that many Native American languages of the Pacific Northwest encode an otherwise rare "yellow-with-green" basic category.⁴²

These examples show that color categorization in its entirety depends upon a tangled hierarchy of perceptual and cognitive processes, some species specific and others culture specific. They also serve to illustrate the point that color categories are not to be found in some pre-given world that is independent of our perceptual and cognitive capacities. The categories red, green, yellow, blue, purple, orange—as well as light/warm, dark/cool, yellow-with-green, etc.—are experiential, consensual, and embodied: they depend upon our biological and cultural history of structural coupling.

We can now appreciate, then, how color provides a paradigm of a cognitive domain that is neither pre-given nor represented but rather experiential and enacted. It is very important to note that just because color is not pre-given does not mean that it does not exhibit universals or that it cannot yield to rigorous analysis by the various branches of science. Since color provides such a paradigm, we will return to it at various points. The time has come, however, to step back and consider some of the lessons this cognitive domain provides for our understanding of perception and cognition in general.

Cognition as Embodied Action

Let us begin, once again, with visual perception. Consider the question, "Which came first, the world or the image?" The answer of most vision research—both cognitivist and connectionist—is unambiguously given by the names of the tasks investigated. Thus researchers speak of "recovering shape from shading," "depth from motion," or "color from varying illuminants." We call this stance the *chicken position*:

Chicken position: The world out there has pre-given properties. These exist prior to the image that is cast on the cognitive system, whose task is to recover them appropriately (whether through symbols or global subsymbolic states).

Notice how very reasonable this position sounds and how difficult it is to imagine that things could be otherwise. We tend to think that the only alternative is the *egg position*:

Egg position: The cognitive system projects its own world, and the apparent reality of this world is merely a reflection of internal laws of the system.

Our discussion of color suggests a middle way between these two chicken and egg extremes. We have seen that colors are not "out there" independent of our perceptual and cognitive capacities. We have also seen that colors are not "in here" independent of our surrounding biological and cultural world. Contrary to the objectivist view, color categories are experiential; contrary to the subjectivist view, color categories belong to our shared biological and cultural world. Thus color as a study case enables us to appreciate the obvious point that chicken and egg, world and perceiver, specify each other.

It is precisely this emphasis on mutual specification that enables us to negotiate a middle path between the Scylla of cognition as the recovery of a pre-given outer world (realism) and the Charybdis of cognition as the projection of a pre-given inner world (idealism). These two extremes both take representation as their central notion: in the first case representation is used to recover what is outer; in the second case it is used to project what is inner. Our intention is to bypass entirely this logical geography of inner versus outer by studying cognition not as recovery or projection but as embodied action.

Let us explain what we mean by this phrase *embodied action*. By using the term *embodied* we mean to highlight two points: first, that cognition depends upon the kinds of experience that come from having a body with various sensorimotor capacities, and second, that these individual sensorimotor capacities are themselves embedded in a more encompassing biological, psychological, and cultural context.⁴³ By using the term *action* we mean to emphasize once again that sensory and motor processes, perception and action, are fundamentally inseparable in lived cognition. Indeed, the two are not merely contingently linked in individuals; they have also evolved together.

We can now give a preliminary formulation of what we mean by *enaction*. In a nutshell, the enactive approach consists of two points: (1) perception consists in perceptually guided action and (2) cognitive structures emerge from the recurrent sensorimotor patterns that enable action to be perceptually guided. These two statements will perhaps appear somewhat opaque, but their meaning will become more transparent as we proceed.

Let us begin with the notion of perceptually guided action. We have already seen that for the representationist the point of departure for understanding perception is the information-processing problem of recovering pre-given properties of the world. In contrast, the point of departure for the enactive approach is the study of how the perceiver can guide his actions in his local situation. Since these local situations constantly change as a result of the perceiver's activity, the reference point for understanding perception is no longer a pre-given, perceiver-independent world but rather the sensorimotor structure of the perceiver (the way in which the nervous system links sensory and motor surfaces). This structure—the manner in which the perceiver is embodied—rather than some pre-given world determines how the perceiver can act and be modulated by environmental events. Thus the overall concern of an enactive approach to perception is not to determine how some perceiver-independent world is to be recovered; it is, rather, to determine the common principles or lawful linkages between sensory and motor systems that explain how action can be perceptually guided in a perceiver-dependent world.⁴⁴

This approach to perception was in fact among the central insights of the analysis undertaken by Merleau-Ponty in his early work. It is therefore worthwhile to quote one of his more visionary passages in full:

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. ... "The properties of the object and the intentions of the subject ... are not only intermingled; they also constitute a new whole." 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 the 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 external influences, one could also say that *behavior is the first cause of all the stimulations*.

Thus the form of the excitant is *created by* the organism itself, by its proper manner of offering itself to actions from the 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 keyboard 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 [italics added].⁴⁵

In such an approach, then, perception is not simply embedded within and constrained by the surrounding world; it also contributes to the enactment of this surrounding world. Thus as Merleau-Ponty notes, the organism both initiates and is shaped by the environment. Merleau-Ponty clearly recognized, then, that we must see the organism and environment as bound together in reciprocal specification and selection.

Let us now provide a few illustrations of the perceptual guidance of action. In a classic study, Held and Hein raised kittens in the dark and exposed them to light only under controlled conditions.⁴⁶ A first group of animals was allowed to move around normally, but each of them was harnessed to a simple carriage and basket that contained a member of the second group of animals. The two groups therefore shared the same visual experience, but the second group was entirely passive. When the animals were released after a few weeks of this treatment, the first group of kittens behaved normally, but those who had been carried around behaved as if they were blind: they bumped into objects and fell over edges. This

beautiful study supports the enactive view that objects are not seen by the visual extraction of features but rather by the visual guidance of action.

Lest the reader feel that this example is fine for cats but removed from human experience, consider another case. Bach y Rita has designed a video camera for blind persons that can stimulate multiple points in the skin by electrically activated vibration.⁴⁷ Using this technique, images formed with the camera were made to correspond to patterns of skin stimulation, thereby substituting for the visual loss. Patterns projected on to the skin have no "visual" content unless the individual is behaviorally active by directing the video camera using head, hand, or body movements. When the blind person does actively behave in this way, after a few hours of experience a remarkable emergence takes place: the person no longer interprets the skin sensations as body related but as images projected into the space being explored by the bodily directed "gaze" of the video camera. Thus to experience "real objects out there," the person must actively direct the camera (by head or hand).

Another sensory modality where the relation between perception and action can be seen is olfaction. Over many years of research, Walter Freeman has managed to insert an array of electrodes into the olfactory bulb of a rabbit so that a small portion of the global activity can be measured while the animal behaves freely.⁴⁸ He found that there is no clear pattern of global activity in the bulb unless the animal is exposed to one specific odor several times. Furthermore, such emergent patterns of activity seem to be created out of a background of incoherent or chaotic activity into a coherent attractor.⁴⁹ As in the case of color, smell is not a passive mapping of external features but a creative form of enacting significance on the basis of the animal's embodied history.

There is in fact growing evidence that this kind of fast dynamics can underlie the configuration of neuronal ensembles. It has been reported in the visual cortex in cats and monkeys linked to visual stimulation; it has been found in radically different neural structures such as the avian brain and even the ganglia of an invertebrate, *Hermisenda*.⁵⁰ This universality is important, for it indicates the fundamental nature of this kind of mechanism of sensorimotor coupling and hence enaction. Had this kind of mechanism been a more species-specific process, typical of, say, only the mammalian cortex, it would have been far less convincing as a working hypothesis.⁵¹

Let us now turn to the idea that cognitive structures emerge from the kinds of recurrent sensorimotor patterns that enable action to be perceptually guided. The pioneer and giant in this area is Jean Piaget.⁵² Piaget laid out a program that he called *genetic epistemology*: he set himself the task of explaining the development of the child from an immature biological organism at birth to a being with abstract reason in adulthood. The child begins with only her sensorimotor system, and Piaget wishes to understand how sensorimotor intelligence evolves into the child's conception of an external world with permanent objects located in space and time and into the child's conception of herself as both an object among other objects and as an internal mind. Within Piaget's system, the newborn infant is neither an objectivist nor an idealist; she has only her own activity, and even the simplest act of recognition of an object can be understood only in terms of her own activity. Out of this, she must construct the entire edifice of the phenomenal world with its laws and logic. This is a clear example in which cognitive structures are shown to emerge from recurrent patterns (in Piaget's language, "circular reactions") of sensorimotor activity.

Piaget, however, as a theorist, never seems to have doubted the existence of a pre-given world and an independent knower with a pre-given logical endpoint for cognitive development. The laws of cognitive development, even at the sensorimotor stage, are an assimilation of and an accommodation to that pre-given world. We thus have an interesting tension in Piaget's work: an objectivist theorist who postulates his subject matter, the child, as an enactive agent, but an enactive agent who evolves inexorably into an objectivist theorist. Piaget's work, already influential in some domains, would bear more attention from non-Piagetians.

One of the most fundamental cognitive activities that all organisms perform is categorization. By this means the uniqueness of each experience is transformed into the more limited set of learned, meaningful categories to which humans and other organisms respond. In the behaviorist era of psychology (which was also the heyday of cultural relativism in anthropology), categories were treated as arbitrary, and categorization tasks were used in psychology only to study the laws of learning.⁵³ (The sense of arbitrariness also reflects the subjectivist trends in contemporary thought that emphasize the element of interpretation in all experience.) In the enactive view, although mind and world arise together in enactment, their manner of arising in any particular situation is not arbitrary. Consider the object on

which you are sitting, and ask yourself what it is. What is its name? If you are sitting on a chair, the chances are that you will have thought *chair* rather than *furniture* or *armchair*. Why? Rosch proposed that there was a basic level of categorization in taxonomies of concrete objects at which biology, culture, and cognitive needs for informativeness and economy all met.⁵⁴ In a series of experiments, Rosch et al. found the basic level of categorization to be the most inclusive level at which category members (1) are used, or interacted with, by similar motor actions, (2) have similar perceived shapes and can be imaged, (3) have identifiable humanly meaningful attributes, (4) are categorized by young children, and (5) have linguistic primacy (in several senses).⁵⁵

The basic level of categorization, thus, appears to be the point at which cognition and environment become simultaneously enacted. The object appears to the perceiver as affording certain kinds of interactions, and the perceiver uses the objects with his body and mind in the afforded manner. Form and function, normally investigated as opposing properties, are aspects of the same process, and organisms are highly sensitive to their coordination. And the activities performed by the perceiver/actor with basic-level objects are part of the cultural, consensually validated forms of the life of the community in which the human and the object are situated—they are basic-level activities.

Mark Johnson proposed another very intriguing basic categorization process.⁵⁶ Humans, he argues, have very general cognitive structures called *kinesthetic image schemas*: for example, the *container schema*, the *part-whole schema*, and the *source-path-goal schema*. These schemas originate in bodily experience, can be defined in terms of certain structural elements, have a basic logic, and can be metaphorically projected to give structure to a wide variety of cognitive domains. Thus, the container schema's structural elements are "interior, boundary, exterior," its basic logic is "inside or outside," and its metaphorical projection gives structure to our conceptualizations of the visual field (things go in and out of sight), personal relationships (one gets in or out of a relationship), the logic of sets (sets contain their members), and so on.

On the basis of a detailed study of these kinds of examples, Johnson argues that image schemas emerge from certain basic forms of sensorimotor activities and interactions and so provide a preconceptual structure to our experience. He argues that since our conceptual understanding is shaped by experience, we also have image-schematic concepts. These concepts have a

basic logic, which imparts structure to the cognitive domains into which they are imaginatively projected. Finally, these projections are not arbitrary but are accomplished through metaphorical and metonymical mapping procedures that are themselves motivated by the structures of bodily experience. Sweetzer provides specific case studies of this process in linguistics. She argues that historical changes of meaning of words in languages can be explained as metaphorical extensions from the concrete and bodily relevant senses of basic-level categories and image schemas to more abstract meanings—for example, “to see” comes to mean “to understand.”⁵⁷

Focusing on categorization, Lakoff has written a compendium of the work that various people have done that can be interpreted to challenge an objectivist viewpoint.⁵⁸ Recently Lakoff and Johnson have produced a manifesto of what they call an experientialist approach to cognition. This is the central theme of their approach:

Meaningful conceptual structures arise from two sources: (1) from the structured nature of bodily and social experience and (2) from our innate capacity to imaginatively project from certain well-structured aspects of bodily and interactional experience to abstract conceptual structures. Rational thought is the application of very general cognitive processes—focusing, scanning, superimposition, figure-ground reversal, etc.—to such structures.⁵⁹

This statement would seem consonant with the view of cognition as enactment for which we are arguing.

One provocative possible extension of the view of cognition as enactment is to the domain of cultural knowledge in anthropology. Where is the locus of cultural knowledge such as folktales, names for fishes, jokes—is it in the mind of the individual? In the rules of society? In cultural artifacts? How can we account for the variation found across time and across informants?⁶⁰ Great leverage for anthropological theory might be obtained by considering the knowledge to be found in the interface between mind, society, and culture rather than in one or even in all of them. The knowledge does not preexist in any one place or form but is enacted in particular situations—when a folktale is told or a fish named. We leave it to anthropology to explore this possibility.

Heideggerian Psychoanalysis

A view of psychopathology fundamentally different from either the Freudian approach or object relations theory was offered by Karl Jaspers, Ludwig

Binswagner, and Merleau-Ponty based on the philosophy of Heidegger.⁶¹ Intended to account for psychological disorders more general, more characterological, than the hysterical and compulsive symptomatology in which Freudian analysis specializes, this account can be dubbed the ontological view to contrast with Freud's representational, cognitivist, epistemological view.⁶² In the ontological view, a character disorder can be understood only in terms of a person's entire mode of being in the world. A theme, such as inferiority and dominance, which is usually only one dimension among many used by an individual in defining his world, becomes fixated, through an early experience, such that it becomes the only mode through which the person can experience himself in the world. It becomes like the light by which objects are seen—the light itself cannot be seen as an object—and thus there is no comparison possible with other modes of being in the world.⁶³ Existential psychoanalysis has extended this type of analysis to pathologies other than character disorders at the same time that it has recharacterized so-called pathologies as existential choices.⁶⁴

The extent to which this phenomenological portrait of pathology lacks any specific methods of its own for treatment is well known, however. The patient might attempt to recall the initial incidents that produced the totalizing of one theme, enact and work through this theme through transference with the therapist, or undergo body work to discover and alleviate the embodied stance of the theme—all, however, are equally characteristic of therapies in which the disorder is conceived in a Freudian, object relations, or other theoretical fashion.

The possibilities for total personal reembodiment inherent in the mindful, open-ended approach to experience that we have been describing may provide the needed framework and tools for implementation of an existential, embodied psychoanalysis. In fact, the relationship between meditation practice, Buddhist teachings, and therapy is a topic of great interest and great controversy among Western mindfulness-awareness practitioners.⁶⁵ Psychological therapy in the Western sense is a historically and culturally unique phenomenon; there is no specific counterpart within traditional Buddhism. Many Western meditators (whether they consider themselves students of Buddhism or not) either are therapists or are considering becoming therapists, and many more have the experience of undergoing therapy. But again, we must remind the reader of our disclaimer concerning what is said in this book about psychoanalysis. An adequate discussion of this

ferment would lead us too far afield at this point, but we invite the reader to consider what form a reembodimenting psychoanalysis might take.

The Retreat into Natural Selection

In preparation for the next chapter, we now wish to take note of a prevalent view within cognitive science, one which constitutes a challenge to the view of cognition that we have presented so far. Consider, then, the following response to our discussion: "I am willing to grant that you have shown that cognition is not simply a matter of representation but depends on our embodied capacities for action. I am also willing to grant that both our perception and categorization of, say, color, are inseparable from our perceptually guided activity and that they are enacted by our history of structural coupling. Nevertheless, this history is not the result of just any pattern of coupling; it is largely the result of biological evolution and its mechanism of natural selection. Therefore our perception and cognition have *survival value*, and so they must provide us with some more or less *optimal fit* to the world. Thus, to use color once more as an example, it is this optimal fit between us and the world that explains why we see the colors we do."

We do not mean to attribute this view to any particular theory within cognitive science. On the contrary, this view can be found virtually anywhere within the field: in vision research, it is common both to the computational theory of Marr and Poggio⁶⁶ and to the "direct theory" of J. J. Gibson and his followers.⁶⁷ It is prevalent in virtually every aspect of the philosophical project of "naturalized epistemology."⁶⁸ It is even voiced by those who insist on an embodied and experientialist approach to cognition.⁶⁹ For this reason, this view can be said to constitute the "received view" within cognitive science of the evolutionary basis for cognition. We cannot ignore, then, this retreat into natural selection.

Let us begin, once again, with our now familiar case study of color. The cooperative neuronal operations underlying our perception of color have resulted from the long biological evolution of the primate group. As we have seen, these operations partly determine the basic color categories that are common to all humans. The prevalence of these categories might lead us to suppose that they are optimal in some evolutionary sense, even though they do not reflect some pre-given world.

This conclusion, however, would be considerably unwarranted. We can safely conclude that since our biological lineage has continued, our color categories are *viable* or *effective*. Other species, however, have evolved *different* perceived worlds of color on the basis of different cooperative neuronal operations. Indeed, it is fair to say that the neuronal processes underlying human color perception are rather peculiar to the primate group. Most vertebrates (fishes, amphibians, and birds) have quite different and intricate color vision mechanisms. Insects have evolved radically different constitutions associated with their compound eyes.⁷⁰

One of the most interesting ways to pursue this comparative investigation is through a comparison of the dimensionalities of color vision. Our color vision is *trichromatic*: as we have seen, our visual system comprises three types of photoreceptors cross-connected to three color channels. Therefore, three dimensions are needed to represent our color vision, that is, the kinds of color distinctions that we can make. Trichromacy is certainly not unique to humans; indeed, it would appear that virtually every animal class contains some species with trichromatic vision. More interesting, however, is that some animals are *dichromats*, others are *tetrachromats*, and some may even be *pentachromats*. (Dichromats include squirrels, rabbits, tree shrews, some fishes, possibly cats, and some New World monkeys; tetrachromats include fishes that live close to the surface of the water like goldfish, and diurnal birds like the pigeon and the duck; diurnal birds may even be pentachromats).⁷¹ Whereas two dimensions are needed to represent dichromatic vision, four are needed for tetrachromatic vision (see figure 8.6), and five for pentachromatic vision. Particularly interesting are tetrachromats (perhaps pentachromats) birds, for their underlying neuronal operations appear to differ dramatically from ours.⁷²

When people hear of this evidence for tetrachromacy, they respond by asking, "What are the other colors that these animals see?" This question is understandable but naive if it is taken to suggest that tetrachromats are simply better at seeing the colors we see. It must be remembered, though, that a four-dimensional color space is fundamentally different from a three-dimensional one: strictly speaking, the two color spaces are incommensurable, for there is no way to map the kinds of distinctions available in four dimensions into the kinds of distinctions available in three dimensions without remainder. We can, of course, obtain some analogical insights into what such higher dimensional color spaces might be like. We could

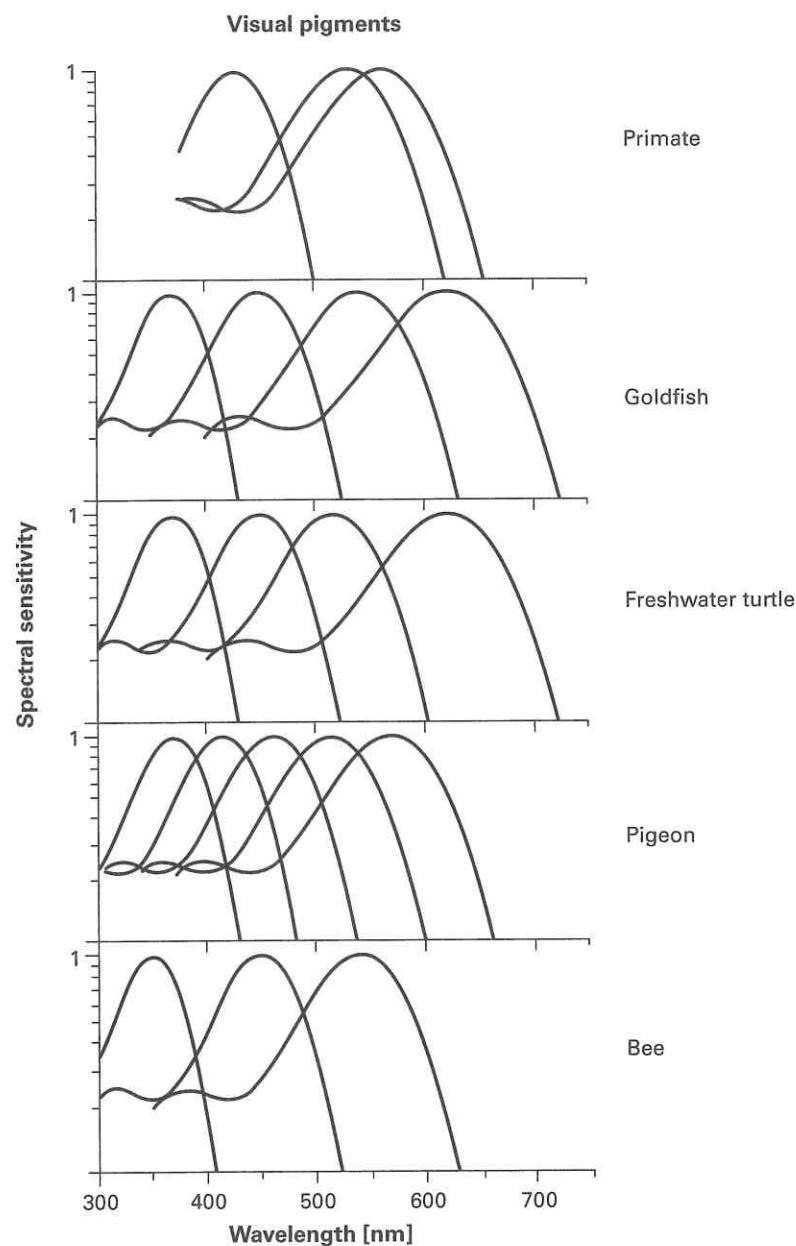


Figure 8.6

Tetrachromatic vs. trichromatic mechanisms are illustrated here on the basis of the different retinal pigments present in various animals. From Neumeyer, *Das Farbensehen des Goldfisches*.

imagine, for example, that our color space contains an additional temporal dimension. In this analogy, colors would flicker to different degrees in proportion to the fourth dimension. Thus to use the term *pink*, for example, as a designator in such a four-dimensional color space would be insufficient to pick out a single color: one would have to say *rapid-pink*, etc. If it turns out that the color space of diurnal birds is pentachromatic (which is indeed possible), then we are simply at a loss to envision what their color experience could be like.⁷³

It should now be apparent, then, that the vastly different histories of structural coupling for birds, fishes, insects, and primates have enacted or brought forth different perceived worlds of color. Therefore, our perceived world of color should not be considered to be the optimal “solution” to some evolutionarily posed “problem.” Our perceived world of color is, rather, a result of one possible and viable phylogenetic pathway among many others realized in the evolutionary history of living beings.

Again, the response on the behalf of the “received view” of evolution in cognitive science will be, “Very well, let us grant that color as an attribute of our perceived world cannot be explained simply by invoking some optimal fit, since there is such a rich diversity of perceived worlds of color. Thus the diverse neuronal mechanisms underlying color perception are not different solutions to the same evolutionarily posed problem. But all that follows is that our analysis must be made more precise. These various perceived worlds of color reflect various forms of adaptation to diverse ecological niches. Each animal group optimally exploits different regularities of the world. It is still a matter of optimal fit with the world; it is just that each animal group has its own optimal fit.”

This response is a still more refined form of the evolutionary argument. Although optimizations are considered to differ according to the species in question, the view remains that perceptual and cognitive tasks involve some form of optimal adaptation to the world. This view represents a sophisticated neorealism, which has the notion of *optimization* as its central explanatory tool. We cannot proceed further, then, without examining more closely this idea in the context of evolutionary explanations. We cannot attempt to summarize the state of the art of evolutionary biology today, but we do need to explore some of its classical foundations and their modern alternatives.

Chapter 7

1. R. Rorty, *Philosophy and the Mirror of Nature*.
2. See Searle, *Intentionality*.
3. This conception of vision is, of course, due to David Marr. See Marr, *Vision*, especially the introduction. For a philosophical explication of the idea of information involved in the representationalist approach, see Dretske, *Knowledge and the Flow of Information*.
4. See Quine, Epistemology naturalized; and the other essays collected in Kornblith, *Naturalizing Epistemology*.
5. See R. Rorty, *Philosophy and the Mirror of Nature*, 246.
6. Fodor, Fodor's guide to mental representations.
7. Minsky, *The Society of Mind*, 287.
8. Ibid., 288. The italics are Minsky's.
9. For a detailed discussion of this notion of operational closure, see Varela, *Principals of Biological Autonomy*.
10. See ibid.; Kelso and Kay, Information and control.
11. Bernstein, *Beyond Objectivism and Relativism*, part III.
12. Kant, *Critique of Pure Reason*, 257.
13. In his replies to Hobbes's objections, Descartes wrote, "I take the term idea to stand for whatever the mind directly perceives. ... I employed this term because it was the term currently used by the Philosophers for the forms of perception of the Divine mind, though we can discern no imagery in God; besides I had no more suitable term." *The Philosophical Works of Descartes, Volume II*, 67–68.
14. See R. Rorty, *Philosophy and the Mirror of Nature*, chapter 1.
15. Minsky, *The Society of Mind*, 304.
16. Ibid.

Chapter 8

1. H. Dreyfus and S. Dreyfus, *Mind over Machine*.
2. See Winograd and Flores, *Understanding Computers and Cognition*. Our argument in this section owes a great deal to this work.
3. For an account of regularization theory, see Poggio, Torre, and Koch, Computational vision and regularization theory.

4. For a sample of discussions in AI about these themes, see the multiple reviews of Winograd and Flores, *Understanding Computers and Cognition*, in *Artificial Intelligence* 31 (1987): 213–261.
5. This point was first made by H. Dreyfus, *What Computers Can't Do*. For a more recent argument to this effect, see Putnam, *Much ado about not very much*.
6. See Heidegger, *Being and Time*; Gadamer, *Truth and Method*. For an introduction to hermeneutics see Palmer, *Hermeneutics*.
7. For references to phenomenology see chapter 2. In this connection, the work of M. Foucault is also essential. See Foucault, *The Order of Things*; Foucault, *Discipline and Punish*. For a critical discussion of Foucault in relation to both hermeneutics and phenomenology, see Dreyfus and Rabinow, *Michel Foucault*.
8. For an exception to this view of folk psychology, one that defends a "first-person" approach in which folk-psychology is not a "third-person" causal-explanatory theory, see Thornton, *Folk Psychology*.
9. Johnson, *The Body in the Mind*, 175.
10. Ibid., 14.
11. This model was first introduced in Varela, Structural coupling and the origin of meaning in a simple cellular automata.
12. For more details, see Varela, *Principles of Biological Autonomy*.
13. Hurvich and Jameson, An opponent-process theory of color vision. For more recent developments, see the articles by Hurvich and Jameson in Ottoson and Zeki, *Central and Peripheral Mechanisms of Colour Vision*.
14. The most recent demonstrations are due to E. Land. See Land, The retinex theory of color vision; and for recent developments, Land, Recent advances in retinex theory and some implications for cortical computations. For earlier discussions, see Helson, Fundamental problems in color vision. I; Helson and Jeffers, Fundamental problems in color vision. II; Judd, Hue, saturation, and lightness of surface colors with chromatic illumination.
15. For a vivid demonstration of these two phenomena, see Brou et al., The colors of things.
16. This experiment belongs to the kind of phenomena made popular by E. Land. See Land, Experiments in color vision; Land, The retinex. The use of rotations of gray checkerboards, as described here, was first presented in Maturana, Uribe, and French, A biological theory of relativistic color coding in the primate retina.
17. See Gouras and Zenner, Color vision.
18. Zeki, Colour coding in the cerebral cortex.

19. Kandinsky, *Concerning the Spiritual in Art*, 57. As quoted in Johnson, *The Body in the Mind*, 83–84.
20. Johnson, *The Body in the Mind*, 84.
21. For an excellent recent review see DeYoe and Van Essen, Concurrent processing streams in monkey visual cortex.
22. Sacks and Wasserman, The case of the colorblind painter.
23. Ibid., 26.
24. Ibid., 33.
25. Maloney, *Computational Approaches to Color Constancy*; Maloney and Wandell, Color constancy; see also Gershon, *The Use of Color in Computational Vision*.
26. See Maloney, *Computational Approaches to Color Constancy*, 119. For a philosophical discussion, see Hilbert, *Color and Color Perception*; Matthen, Biological functions and perceptual content. For extensive discussion and criticism of this view, see Thompson, *Colour Vision*.
27. For detailed arguments, see Hardin, *Color for Philosophers*; Thompson, *Colour Vision*.
28. See Jameson and Hurvich, Essay concerning color constancy.
29. Gouras and Zenner, Color vision, 172.
30. Consider, for example, this passage from a well-known text by Gleason, *An Introduction to Descriptive Linguistics*, 4: "There is a continuous gradation of color from one end of the spectrum to the other. Yet an American describing it will list the hues as red, orange, yellow, green, blue, purple, or something of the kind. There is nothing inherent either in the spectrum or the human perception of it which would compel its division in this way."
31. Berlin and Kay, *Basic Color Terms*.
32. Ibid., 109.
33. E.R. Heider [Rosch], Universals in color naming and memory.
34. Brown and Lenneberg, A study in language and cognition; Lantz and Steffire, Language and cognition revisited; Steffire, Castillo Vales, and Morely, Language and cognition in yucatan.
35. Heider [Rosch], Universals in color naming and memory; Heider [Rosch], Linguistic relativity; Rosch, On the internal structure of perceptual and semantic categories; Heider [Rosch] and Olivier, The structure of the color space in naming and memory for two languages.

36. Heider [Rosch], Focal color areas and the development of color names.
37. Lakoff, *Women, Fire and Dangerous Things*.
38. Kay and McDaniel, The linguistic significance of the meanings of basic color terms.
39. DeValois and Jacobs, Primate color vision.
40. Kay and Kempton, What is the Sapir-Whorf hypothesis?
41. As reported in Lakoff, *Women, Fire and Dangerous Things*, 29.
42. MacLaury, Color-category evolution and Shuswap yellow-with-green.
43. This conception of embodiment has been most emphasized in cognitive science by H. Dreyfus, *What Computers Can't Do*; Johnson, *The Body in the Mind*; and Lakoff, *Women, Fire and Dangerous Things*.
44. See Kelso and Kay, Information and control.
45. Merleau-Ponty, *The Structure of Behavior*, 13.
46. Held and Hein, Adaptation of disarranged hand-eye coordination contingent upon re-afferent stimulation.
47. Bach y Rita, *Brain Mechanisms in Sensory Substitution*, as described in Livingstone, *Sensory Processing, Perception, and Behavior*.
48. Freeman, *Mass Action in the Nervous System*.
49. Freeman and Skarda, Spatial EEG patterns, nonlinear dynamics, and perception.
50. For a recent review see Bressler, The gamma wave; the work of Gray and Singer, Stimulus-specific neuronal oscillations in orientation columns in cat visual cortex, has been largely responsible for the wider acceptance of this hypothesis; for Hermisenda see Gelperin and Tank, Odour-modulated collective network oscillations of olfactory interneurons in a terrestrial mollusc; and for the results on the avian brain see Neuenschwander and Varela, Sensori-triggered and spontaneous oscillations in the avian brain.
51. It should also be noted that this fast dynamics is not restricted to sensorial trigger: the oscillations appear and disappear quickly and quite spontaneously in various places of the brain. This suggests that such fast dynamics involve all those subnetworks that give rise to the entire readiness-to-hand in the next moment. They involve not only sensory interpretation and motor action but also the entire gamut of cognitive expectations and emotional tonality, which are central to the shaping of a moment of action. Between breakdown these oscillations are the symptoms of (rapid) reciprocal cooperation and competition among distinct agents that are activated by the current situation, vying with each other for differing modes of

interpretation for a coherent cognitive framework and readiness for action. On the basis of this fast dynamics, as in an evolutionary process, one neuronal ensemble (one cognitive subnetwork) finally becomes more prevalent and becomes the behavioral mode for the next cognitive moment. When we say "becomes prevalent," we do not mean a process of optimization but rather a process of consolidation out of a chaotic dynamic.

52. All of Piaget's books are relevant. We are particularly indebted to Piaget, *The Construction of Reality in the Child*.

53. See for example Bourne, Dominowski, and Loftus, *Cognitive Processes*.

54. E. Rosch et al., Basic objects in natural categories; Rosch, Principles of categorization; Rosch, Wittgenstein and categorization research in cognitive psychology; Mervis and Rosch, Categorization of natural objects.

55. Rosch et al., Basic objects in natural categories.

56. Johnson, *The Body in the Mind*.

57. Sweetzer, *Semantic Structure and Semantic Change*.

58. Lakoff, *Women, Fire and Dangerous Things*.

59. Lakoff, Cognitive semantics. This article provides a concise overview of Lakoff and Johnson's experientialist approach.

60. Berofski, *Making History*.

61. Merleau-Ponty, *Phenomenology of Perception*; Jaspers, *Allgemeine psychopathologie*; Binswanger, *Zur phänomenologischen Anthropologie*.

62. H. Dreyfus, Alternative philosophical conceptualizations of psychopathology.

63. This is reminiscent of the Buddhist view that consciousness is always born into a total realm. See chapter 4, note 12.

64. The classic statement is May, *Existential Psychoanalysis*.

65. Wilber, Engler, and Brown, *Transformations of Consciousness*; Wellwood, *Awakening the Heart*.

66. Marr, *Vision*; Poggio, Torre, and Koch, Computational vision and regularization theory.

67. Gibson, *The Ecological Approach to Visual Perception*.

68. Kornblith, *Naturalizing Epistemology*.

69. This tendency can occasionally be discerned in both Lakoff, *Women, Fire and Dangerous Things* and Johnson, *The Body in the Mind*.

70. For comparative discussions of color vision, see Jacobs, *Comparative Color Vision*; Nuboer, A comparative review on colour vision. For insect color vision, see Menzel, Spectral sensitivity and colour vision in invertebrates. For discussion in the context of cognitive science, see Thompson, Palacios, and Varela, Ways of coloring.

71. For tetrachromacy in fishes, see Harosi and Hashimoto, Ultraviolet visual pigment in a vertebrate; Neumeyer, *Das Farbensehen des Goldfisches*. For birds, see Jane and Bowmaker, Tetrachromatic colour vision in the duck; Burkhardt, UV vision; Palacios et al., Color mixing in the pigeon; Palacios and Varela, Color mixing in the pigeon. II.

72. These mechanisms have still not been studied with the detail that has been devoted to those in the primate group. See Varela et al., The neurophysiology of avian color vision.

73. For extensive discussion of these and other implications of comparative color vision in a philosophical context, see Thompson, *Colour Vision*; and Thompson et al., Ways of coloring.

Chapter 9

1. See in particular Gould, Darwinism and the expansion of evolutionary theory; Gould and Lewontin, The spandrels of San Marco and the Panglossian paradigm. For more general discussion, see Sober, *The Nature of Selection*; Ho and Saunders, *Beyond Neo-Darwinism*; Endler, The newer synthesis? For a recent defense of neo-Darwinism in the face of these various challenges see Hecht and Hoffman, Why not neo-Darwinism? Piatelli-Palmarini, Evolution, selection, and cognition, explores similar themes, though in the context of a defense of cognitivism.

2. This term is from Sober, *The Nature of Selection*.

3. The idea of evolution as natural drift was first introduced in Maturana and Varela, *The Tree of Knowledge*. In this chapter we expand and modify this idea significantly in relation to its original presentation.

4. Geschwind and Galaburda, *Cerebral Lateralization*.

5. Gould and Eldredge, Punctuated equilibria.

6. Packard, An intrinsic model of adaptation.

7. For a concise comparison between these two extremes see Lambert and Hughes, Keywords and concepts in structuralist and functionalist biology.

8. For this topic see the articles in Goodwin, Holder, and Wyles, *Development and Evolution*.

9. de Beer, *Embryos and Ancestors*, 163.