THE SECOND WAVE OF CYBERNETICS: FROM REFLEXIVITY TO SELF-ORGANIZATION

It all started with a frog. In a classic article entitled "What the Frog's Eye Tells the Frog's Brain," central players in the Macy group—including Warren McCulloch, Walter Pitts, and Jerry Lettvin—did pioneering work on a frog's visual system. They demonstrated, with great elegance, that the frog's visual system does not so much *represent* reality as *construct* it.[1] What's true for frogs must also hold for humans, for there's no reason to believe that the human neural system is uniquely constructed to show the world as it "really" is. Not everyone in the research group was interested in pursuing the potentially radical epistemological implications of this work. McCulloch, for example, remained wedded to realist epistemology. But a young neurophysiologist from Chile, Humberto Maturana, was also on the research team, and he used it as a springboard into the unknown. Pushing the envelope of traditional scientific objectivity, he developed a new way of talking about life and about the observer's role in describing living systems. Entwined with the epistemological revolution he started are the three stories we have been following: the reification of information, the cultural and technological construction of the cyborg, and the transformation of the human into the posthuman. As a result of work by Maturana and his collaborator, Francisco Varela, all three stories took decisive turns during the second wave of cybernetics, from 1960 to 1985. This chapter follows the paths that Maturana and Varela took as they probed deeply into what it means to acknowledge that the observer, like the frog, does not so much discern preexisting systems as create them through the very act of observation.

Central to the seriated changes connecting these second-wave developments to the first wave is the difficult and protean concept of reflexivity. As we saw in chapter 3, participants in the Macy Conferences wrestled with

reflexivity, without much success. The particularities of the situation—the embedding of reflexivity within psychoanalytic discourse, Kubie's halitosis of the personality, the unquantifiability of reflexive concepts—put a spin on reflexivity that affected its subsequent development.[2] Gregory Bateson's 1968 conference had made clear that the problems posed by including the observer could be addressed only if a substantial reworking of realist epistemology was undertaken. The intuitive leap made by Bateson in concluding that the internal world of subjective experience is a metaphor for the external world remained a flash of insight rather than a quantitatively reliable inference that experimentalists like Warren McCulloch could endorse. The problem was how to make the new epistemology operational by integrating it with an experimental program that would replace intuition with empirical data.

At issue in this evolving series of events are questions crucially important to the technoscientific concepts of information, the cyborg, and the posthuman. Like Norbert Wiener, Maturana has strong ties with liberal humanism. At stake for him was how to preserve the central features of autonomy and individuality while still wrenching them out of the Cartesian and Enlightenment frameworks in which they are embedded. Even as he struggled mightily to "say something new," his work replicates some assumptions of the first wave at the same time that it radically revises others.[3] We can see an early form of the struggle in the essays of Heinz von Foerster, the genial and well-connected Austrian émigré who functions as a transitional figure linking first- and second-wave cybernetics. From this beginning, we will trace the epistemological revolution that Maturana fomented, delineate its connections with the three stories we have been following, and finally explore the differing assumptions that led Varela, Maturana's collaborator, to set off in a new direction.

Reflexivity Revisited

Von Foerster left Austria in 1948, after working on microwave electronics for Germany during World War II, work that had important applications in radar (his 1949 vita lists much of this research as "secret").[4] In the spring of 1949 von Foerster wrote McCulloch, renowned for his generosity in helping younger men, to seek his help in finding a job in North America.[5] McCulloch found the Austrian a position at the University of Illinois; he also introduced von Foerster into the Macy group. Soon afterward, McCulloch and Mead asked von Foerster if he would serve as principal editor of the published transcripts. Although he had some misgivings because

English was not his native language, he agreed. With his name emblazoned on the title pages, the published transcripts are associated with him as much as with anyone.

It was not until the Macy Conferences had run their course, however, that von Foerster tried to develop more fully the epistemological implications of including the observer as part of the system. The punning title of his essay collection, *Observing Systems*, announces reflexivity as a central theme. "Observing" is what (human) systems do; in another sense, (human) systems themselves can be observed. The earliest essay ("On Self-Organizing Systems and Their Environments"), taken from a presentation given in 1960, shows von Foerster thinking about reflexivity as a circular dynamic that can be used to solve the problem of solipsism. How does he know other people exist, he asks. Because he experiences them in his imagination. His experience leads him to believe that other people similarly experience him in their imaginations. "If I assume that I am the sole reality, it turns out that I am the imagination of somebody else, who in turn assumes that *he* is the sole reality."[6] In a circle of intersecting solipsisms, I use my imagination to conceive of someone else and then of the imagination of that person, in which I find myself reflected.[7] Thus I am reassured not only of the other person's existence but of my own as well. Although charmingly posed, the argument is logically nonsensical, for there is no assurance that other imaginations are conceiving of me any more than I am conceiving of them. Maybe I am thinking not about von Foerster but about a Big Mac. That even a fledgling philosopher could reduce the argument to shreds is perhaps beside the point. Von Foerster himself seemed to recognize that the argument was the philosophical equivalent to pulling a rabbit from a hat, for he finally "solves" the paradoxical circling between solipsistic imaginations by asserting what he was to prove, namely the existence of reality.

Although the argument is far from rigorous, it is interesting for the line of thought it suggests. Its implications are illustrated by a cartoon (drawn by Gordon Pask at von Foerster's request) of a man in a bowler hat, in whose head is pictured another man in a bowler hat, in whose head is yet another man in a bowler hat.[8] The potentially infinite regress of men in bowler hats does more than create an image of the observer who observes himself by observing another. It also visually distinguishes the observer as a discrete system inside the larger system of the organism. In the aftermath of the Macy Conferences, one of the central problems with reflexivity was how to talk about it without falling into solipsism or resorting to psychoanalysis. The message from the Macy Conferences was clear: if reflexivity was to be credible, it had to be insulated against subjectivity and presented in a

context in which it had at least the potential for rigorous (preferably mathematical) formulation. As Norbert Wiener was later to proclaim, "Cybernetics is nothing if it is not mathematical."[9] Distinguishing the observer as a system separate from the organism was one way to make reflexivity more manageable, for it reduced the problem of the observer to a problem of communication among systems.

Throughout the 1960s, von Foerster remained convinced of the importance of reflexivity, and he experimented with various ways to formulate it. A breakthrough occurred in 1969, when he invited Maturana to speak at a conference at the University of Illinois. There Maturana unveiled his ideas about treating "cognition as a biological phenomenon."[10] The power of Maturana's theory must have deeply affected von Foerster, for his thinking about reflexivity takes a quantum leap in complexity after this date. The increased sophistication can be seen in his 1970 essay "Molecular Ethology: An Immodest Proposal for Semantic Clarification," in which he criticizes behaviorism by making the reflexive move of turning the focus from the observation back onto the observer. Behaviorism does not demonstrate that animals are black boxes that give predictable outputs for given inputs, he argues. Rather, behaviorism shows the cleverness and power of the experimenter in getting animals to behave as such. "Instead of searching for mechanisms in the environment that turn organisms into trivial machines, we have to find the mechanisms within the organisms that enable them to turn their environment into a trivial machine."[11] Here reflexivity moves from men in bowler hats to the beginning of a powerful critique of objectivist epistemology. By 1972, von Foerster had been so thoroughly convinced by Maturana's theory that one of the latest essays in *Observing Systems,* "Notes on an Epistemology of Living Things" (pp. 258–71), recasts the theory in the form of a circular set of numbered quasi-mathematical propositions, in which the last repeats the first.

To trace the evolution of Maturana's epistemology, let us turn now to the seminal paper "What the Frog's Eye Tells the Frog's Brain." In it, Maturana and his coauthors demonstrate that the frog's sensory receptors speak to the brain in a language highly processed and species-specific. To arrive at this conclusion, the authors implanted microelectrodes in a frog's visual cortex to measure the strength of neural responses to various stimuli. At this point the frog's brain became part of a cybernetic circuit, a bioapparatus reconfigured to produce scientific knowledge. Strictly speaking, the frog's brain had ceased to belong to the frog alone. I will therefore drop the possessive and follow the authors by referring to the frog's brain simply as "the brain" (a phrase that eerily echoes the title of Norbert Wiener's short

story discussed in chapter 5). From the wired-up brain, the researchers discovered that small objects in fast, erratic motion elicited maximum response, whereas large, slow-moving objects evoked little or no response. It is easy to see how such perceptual equipment is adaptive from the frog's point of view, because it allows the frog to perceive flies while ignoring other phenomena irrelevant to its interests. The results implied that the frog's perceptual system does not so much register reality as *construct* it. As the authors noted, their work "shows that the [frog's] eye speaks to the brain in a language already highly organized and interpreted instead of transmitting some more or less accurate copy of the distribution of light upon the receptors."[12] The work led Maturana to the maxim fundamental to his epistemology: "Everything said is said by an observer" (*AC*, p. xxii). No wonder the article was quickly recognized as a classic, for it blew a frog-sized hole in realist epistemology.

Despite the potentially radical implications of the article's *content*, however, its *form* reinscribed conventional realist assumptions of scientific discourse. The results are reported in an objectivist rhetoric that masks the fact they are interpreted through the sensory and cognitive interfaces of embodied researchers, whose perceptions were at least as transformative as the frog's. Years later, Maturana would recall that he and Lettvin continued to work in an objectivist framework even as that framework was being called into question by their research. In the preface to *Autopoiesis and Cognition: The Realization of the Living*, Maturana recalled: "When Jerry Y. Lettvin and I wrote our several articles on frog vision . . . we did it with the implicit assumption that we were handling a clearly defined cognitive situation: there was an objective (absolute) reality, external to the animal, and independent of it (not determined by it), which it could perceive (cognize). . . . But even there the epistemology that guided our thinking and writing was that of an objective reality independent of the observer" (*AC*, p. xiv). Faced with this inconsistency, Maturana had a choice. He could continue to work within the prevailing assumptions of scientific objectivity, or he could devise a new epistemology that would construct a worldview consistent with what he thought the experimental work showed.

The break came with his work on color vision in other animals, including birds and primates. He and his coauthors (not the Macy group this time) found they could not map the visible world of color onto the activity of the nervous system.[13] There was no one-to-one correlation between perception and the world. They could, however, correlate activity in an animal's retina with its *experience* of color. If we think of sense receptors as constituting a boundary between outside and inside, this implies that organiza-

tionally, the retina matches up with the inside, not the outside. From this and other studies, Maturana concluded that perception is not fundamentally representational. He argued that to speak of an objectively existing world is misleading, for the very idea of a world implies a realm that preexists its construction by an observer. Certainly there is something "out there," which for lack of a better term we can call "reality." But it comes into existence for us, and for all living creatures, *only through interactive processes determined solely by the organism's own organization.* "No description of an absolute reality is possible," he and Varela wrote in *Autopoiesis and Cognition,* for such a description "would require an interaction with the absolute to be described, but the representation that would arise from such an interaction would necessarily be determined by the autopoietic organization of the observer. . . hence, the cognitive reality that it would generate would unavoidably be relative to the observer" (*AC,* p. 121). Thus he was led to a premise fundamental to his theory: living systems operate within the boundaries of an organization that closes in on itself and leaves the world on the outside.

With Varela, Maturana developed the implications of this insight in *Autopoiesis and Cognition.* He arrived at his theory, he explains in the introduction, by deciding to treat "the activity of the nervous system as determined by the nervous system itself, and not by the external world; thus the external world would have only a triggering role in the release of the internally-determined activity of the nervous system" (*AC,* p. xv). His key insight was to realize that if the action of the nervous system is determined by its organization, the result is a circular, self-reflexive dynamic. A living system's organization causes certain products to be produced, for example, nucleic acids. These products in turn produce the organization characteristic of that living system. To describe this circularity, he coined the term *autopoiesis* or self-making. "It is the circularity of its organization that makes a living system a unit of interactions," he and Varela wrote in *Autopoiesis and Cognition,* "and it is this circularity that it must maintain in order to remain a living system and to retain its identity through different interactions" (*AC,* p. 9). Building on this premise of autopoietic closure, Maturana developed a new and startlingly different account of how we know the world.[14]

What is this account? One path into it is to regard the account as an attempt to counteract anthropomorphic projection by clearly distinguishing between two domains of description. On the one hand, there is what one can say about the circularity of autopoietic processes in themselves, taking care not to attribute to them anything other than what they exhibit. On the

other hand, there are the inferences that observers draw when they place an autopoietic system in the context of an environment. Seeing system and medium together over a period of time, observers draw connections between cause and effect, past and future. But these are the observers' inferences; they are not intrinsic to the autopoietic processes in themselves. Let's say I see a blue jay flash through the trees and settle on the birdbath. I may think, "Oh, it's getting a drink." Other species, for example those lacking color vision, would react to this triggering event with different constructions. A frog might notice the quick, erratic flight but be oblivious to the blue jay at rest. Each living system thus constructs its environment through the "domain of interactions" made possible by its autopoietic organization. What lies outside that domain does not exist for that system. Maturana, realizing that he was fighting a long tradition of realist assumptions deeply embedded in everyday language, developed an elaborate vocabulary as a prophylactic against having anthropomorphism creep back in. The necessity of finding a new language in which to express his theory was borne home to him during the student revolution in Chile in May 1968. It was then, he wrote in *Autopoiesis and Cognition,* that he discovered that "language was a trap, but the whole experience was a wonderful school in which one could discover how mute, deaf and blind one was . . . one began to listen and one's language began to change; and then, but only then, new things could be said" (*AC,* p. xvi).

Shortly we will analyze places where Maturana, like the participants in the Macy Conferences, seems unable to escape from the tar baby of self-reflexive language. For the moment, however, let us explore the "new things" he tried to say. No doubt the cumbersome—many would not hesitate to call it tortured—quality of his language will be immediately apparent to the reader.[15] Before we judge it harshly, however, we should remember that Maturana was attempting nothing less than to give a different account of how we know the world. Since it is partly through language that humans bring worlds into being for themselves, he was in the impossible position of pulling himself up by his own bootstraps, trying to articulate the new by using the only language available, the *lingua franca* whose meanings had long ago settled along lines very different from those he was trying to envision.

We can start with that most problematic of constructions, the observer. From Maturana's point of view, the "fundamental cognitive operation that an observer performs is the operation of distinction" (*AC,* p. xxii). Influenced by G. Spencer-Brown, Maturana (and even more so Varela in his work *Principles of Biological Autonomy*)[16] sees the operation of distinction

as marking space so that an undifferentiated mass is separated into an inside and an outside or, in Maturana's terminology, into a unity and a medium in which the unity is embedded. Unities distinguished by the observer can be of two types, simple and composite. A simple unity "only has the properties with which it is endowed by the operations of distinction through which it becomes separated from a background." Composite unities, by contrast, have "structure and organization," (*AC*, p. xx), terms that Maturana uses in special senses and that require further explanation.

A composite unity's organization is the complex web of all possible relationships that can be realized by the autopoietic processes as they interact with one another. When Maturana speaks of a system's organization, he does not mean how this web of relationships might be described in abstract form. Rather, he intends *organization* to denote the relations actually instantiated by the autopoietic unity's circular processes. Structure, by contrast, is the particular instantiation that a composite unity enacts at a particular moment. For example, when a female human is born, she has one kind of structure; when she enters puberty, she has another; if she contracts a disease, she has still another. But throughout her lifetime, her organization remains the same: that which is characteristic of a living human. Only when death occurs does her organization change. According to Maturana, this ability of living organisms to conserve their autopoietic organization is the necessary and sufficient condition for them to count as living systems. All living systems are autopoietic, and all physical systems, if autopoietic, can be said to be living (*AC*, p. 82). Thus life and autopoiesis are coextensive with one another. Here's how that proposition sounds in Maturana's terminology. "The living organization is a circular organization which secures the production or maintenance of the components that specify it in such a manner that the product of their functioning is the very same organization that produces them" (*AC*, p. 48).

To account for a system's embeddedness in an environment, Maturana uses the concept of structural coupling. All living organisms must be structurally coupled to their environments to continue living; humans, for example, have to breathe air, drink water, eat food (*AC*, pp. x–xi). In addition, systems may be structurally coupled to each other. For example, a cell within my body may be considered as a system in itself, but it relies for its continued existence on its structural coupling to my body as a whole. Here again the role of the observer becomes important, for Maturana is careful to distinguish between the triggering effect that an event in the medium has on a system structurally coupled with it and the causal relationship that observers construct in their mind when they perceive the system interact-

ing with the environment. When my bird dog sees a pigeon, I may think, "Oh, he's pointing because he sees the bird." But in Maturana's terms, this is an inference I draw from my position in the "descriptive domain" of a human observer (*AC*, p. 121). From the viewpoint of the autopoietic processes, there is only the circular interplay of the processes as they continue to realize their autopoiesis, always operating in the present moment and always producing the organization that also produces them. Thus, time and causality are not intrinsic to the processes themselves but are concepts inferred by an observer. "The present is the time interval necessary for an interaction to take place," Maturana and Varela wrote. "Past, future and time exist only for the observer" (*AC*, p. 18).

Information, coding, and teleology are likewise inferences drawn by an observer rather than qualities intrinsic to autopoietic processes. In the autopoietic account, there are no messages circulating in feedback loops, nor are there even any genetic codes. These are abstractions invented by the observer to explain what is seen; they exist in the observer's "domain of interactions" rather than in autopoiesis itself. "The genetic and nervous system are said to code information about the environment and to represent it in their functional organization. This is untenable," Maturana and Varela noted. "The genetic and nervous systems code processes that specify series of transformations from initial states, which can be decoded only through their actual implementation, not *descriptions* that the observer makes of an environment which lies exclusively in *his* cognitive domain" (*AC*, p. 53). Similarly, "the notion of information refers to the observer's degree of uncertainty in his behavior within a domain of alternatives defined by him, hence the notion of information only applies within his cognitive domain" (*AC*, p. 54). The same applies to teleology. "A living system is not a goal-directed system; it is, like the nervous system, a stable state-determined and strictly deterministic system closed on itself and modulated by interactions not specified by its conduct. These modulations, however, are apparent as modulations only for the observer who beholds the organism or the nervous system externally, from his own conceptual *(descriptive)* perspective, as lying in an environment and as elements in his domain of interactions" (*AC*, p. 50).

One implication of letting go of causality is that systems always behave as they should, which is to say, they always operate in accord with their structures, whatever those may be. In Maturana's world, my car always works, whether it starts or not, because it operates only and always in accord with its structure at the moment. It is I, as an observer, who decides that my car is not working because it will not start. Such "punctuations," as Maturana

and Varela call them, belong to the "domain of the observer" (*AC*, pp. 55–56). Because they are extrinsic to the autopoietic processes, they are also extrinsic to the biological description that Maturana aims to give of life and cognition. In an important essay entitled "Biology of Language," Maturana remarks that the "operation of a structure-determined system is necessarily perfect: that is, it follows a course determined only by neighborhood relations in its structure and by nothing else. It is only in a referential domain, such as the domain of behavior, that an observer can claim that an error has occurred when his or her expectations are not fulfilled."[17]

To assess the changes that the autopoietic view entails, let us turn now to compare its account of living systems with that given by first-wave cybernetics. A convenient focal point for the comparison of the two theories is liberal humanism, where their implications for the construction of subjectivity will become apparent. Having traced these implications, we will then consider the impact of second-wave cybernetics on the entwined stories we have been following: the reification of information, the construction of the cyborg, and the transformation of the human into the posthuman.

Reconfiguring the Liberal Humanist Subject

As we saw in chapter 4, Norbert Wiener had a complex relation to the liberal humanist subject. Father of a theory that put humans and machines into the same category, he was nevertheless committed to creating a cybernetics that would preserve autonomy and individuality. His nightmare was the human reduced to a cog in a rigid machine, losing the flexibility and autonomous functioning that Wiener regarded as the birthright of a cybernetic organism. Echoes of this cybernetic tradition linger in Maturana's description of composite unities as "autopoietic machines" (*AC*, p. 82). Fully aware of the implications of calling autopoietic systems "machines," Maturana makes clear that there is nothing in his theory to prohibit artificial systems from becoming autopoietic unities. "If living systems were machines, they could be made by man," he and Varela point out (*AC*, p. 83). They pooh-pooh the idea that life cannot or should not be created by humans. "There seems to be an intimate fear that the awe with respect to life and the living would disappear if a living system could be not only reproduced, but designed by man. This is nonsense. The beauty of life is not a gift of its inaccessibility to our understanding" (*AC*, p. 83). When Maturana objects to first-wave projects that attributed biological properties to machines, his criticism addresses how life is defined, not the idea that machines can be alive. For example, he criticizes John von Neumann's pro-

posal to create a self-reproducing machine by arguing that von Neumann modeled descriptions that biologists had made rather than autopoietic processes in themselves. Von Neumann modeled inferences about "what appeared to take place in the cell in terms of information content, program and coding. By modeling the processes expressed in these descriptions he produced a machine that could make another machine but he did not model the phenomena of cellular reproduction, heredity and genetics as they take place in living systems."[18]

This critique points to an important change between Maturana's position and that announced by Wiener and his coauthors in their cybernetic manifesto. Whereas the latter argued that it is the system's *behavior* that counts, Maturana argues that it is the *autopoietic processes generating behavior* that count. As we have seen, first-wave researchers concentrated on building artifacts that would behave as cybernetic mechanisms: John von Neumann's self-reproducing machines; Claude Shannon's electronic rat; Ross Ashby's homeostat. By contrast, Maturana and others in the second wave look to systems instantiating processes that count as autopoietic. The homeostat might behave cybernetically, for example, but it does not count as an autopoietic machine because it does not produce the components that produce its organization. Perhaps because of this emphasis on process, autopoietic theory has proven readily adaptable to the analysis of social systems. In autopoietic theory, the machine of interest is much more likely to be the state than Robocop or Terminator.[19]

In first-wave cybernetics, questions of boundary formation were crucial to its constructions of subjectivity. Boundary questions are also important in autopoietic theory. Wiener's anxieties recirculate in discussions about what happens when one autopoietic unity is encapsulated within the boundaries of a larger autopoietic unity, for example when a cell functions as part of a larger machine. Can the cell continue to function as an autonomous entity, or must its functioning be subordinated to the larger unity? To distinguish these two cases, Maturana introduces the term *allopoietic*. Whereas autopoietic unities have as their only goal the continuing production of their autopoiesis, allopoietic unities have as their goal something other than producing their organization. When I drive my car, its functioning is subordinated to the goals I set for it. Instead of the pistons using their energy to repair themselves, for example, they use their energy to turn the drive shaft so that I can get to the store. I function autopoietically, but the car functions allopoietically.

We saw in chapter 4 that cybernetic boundary questions often involve deep ethical and psychological issues, such as those that troubled Wiener

when he envisioned the dissolution of the autonomous liberal subject. In autopoietic theory, one of the principal effects of autopoiesis is to secure for a living system the crucial qualities of autonomy and individuality. Consequently, boundary issues are often played out in discussions of how much autonomy autopoietic systems will retain for themselves and how much autonomy they will demand from the systems with which they are structurally coupled. The distinction between allopoietic and autopoietic gives Maturana a way to talk about power struggles within society. In autopoietic theory, the idea corresponding to Wiener's horror at a man being forced to act as a cog in a machine is a system that is capable of autopoiesis being forced instead to function allopoietically, especially for humans. Maturana's ideal is a human society in which one would "see all human beings as equivalent to oneself, and to love them . . . without demanding from them a larger surrender of individuality and autonomy than the measure that one is willing to accept for oneself while integrating it as an observer" (*AC*, p. xxix). Such a society, he adds, "is in its essence an anarchist society, a society made for and by observers that would not surrender their condition of observers as their only claim to social freedom and mutual respect" (*AC*, p. xxx). In such rhetoric, we can easily hear a reinscription of liberal humanist values, even though the epistemology that Maturana advocates is very different from that which gave rise to the Enlightenment subject.

Yet it would be a mistake to think that Maturana's radicalism can be so easily recuperated back into liberal subjectivity. The split between his position and liberal philosophy becomes obvious when questions of objectivity arise. Consider, for example, his insistence that ethics cannot be separated from scientific inquiry. Instead of accepting the proposition that the scientist simply reports what he or she sees and in this sense remains aloof from ethical considerations, Maturana envisions autopoietic theory as a way to reconnect ethics and science. Emphasizing that autonomy always takes place in the context of structural coupling, autopoiesis rejects the objectivism that drives a wedge between the scientist-observer and the world being observed. For Maturana, observation does not mean that the observer remains separate from what is being observed; on the contrary, the observer can observe only because the observer is structurally coupled to the phenomenon she sees. Expanded to social ethics, this implies "in man as a social being . . . all actions, however individual as expressions of preferences or rejections, constitutively affect the lives of other human beings and, hence, have ethical significance." Structural coupling requires that human beings "as components of a society, necessarily realize their individ-

ual worlds and contribute to the determination of the individual worlds of others" (*AC*, p. xxvi).

Although Maturana thus follows in the liberal tradition of cyberneticians like Wiener in placing a high value on the autonomous individual, the meaning of autonomy has undergone significant change. Autonomy as Maturana envisions it is not consistent with laissez-faire capitalism; it is not consistent with the idea that each person is out for himself and devil take the hindmost; and it is not consistent with the ethical position that a scientist could undertake a research program without being concerned about how the results of the research would be used. In these respects, the individualism and autonomy that Maturana champions challenge the premises embodied in liberal subjectivity at least as much as they reinscribe those premises.

To explore further how liberal subjectivity is both contested and reinscribed in autopoietic theory, let us turn now to Maturana's account of the observer. Nowhere does Maturana depart more clearly from first-wave philosophies than in his insistence that the observer must be taken into account. "*The observer is a living system and any understanding of cognition as a biological phenomenon must account for the observer and his role in it*" (*AC*, p. 48). The act of observation necessarily entails reflexivity, for one of the systems that an observer can describe is the observer as an autopoietic system. But reflexivity as Maturana envisions it is very different from the psychoanalytic reflexivity that Lawrence Kubie introduced into the Macy Conferences (see chapter 3). In contrast to Kubie's emphasis on unconscious symbolism, Maturana's observer does not have psychological depth or specificity. Rather, Maturana's observer is more like the observer that Albert Einstein posits in the special theory of relativity. The one who sees is always called simply "the observer," without further specification, implying that any individual of that species occupying that position would see more or less the same thing. Although the observer's perceptions construct reality rather than passively perceive it, for Maturana this construction depends on *positionality* rather than *personality*. In autopoietic theory, the opposite of objectivism is not subjectivism but relativism.

If the interplay between conscious and unconscious processes is not important for Maturana, how is the observer produced? The observer begins as an autopoietic unity, as all living systems are said to be. As a particular kind of autopoietic unity capable of becoming an observer, the observer-system can generate representations of its own interactions. When the system recursively interacts with these representations, it becomes an observer. The system can then recursively generate representations of these

representations and interact with them, as when an observer thinks, "I am an observing system observing itself observing." Each twist of this reflexive spiral adds additional complexity, enlarging the domain of interactions that specify the world for that autopoietic unity. Maturana and Varela summarize the situation thus in *Autopoiesis and Cognition:* "We become *observers* through recursively generating representations of our interactions, and by interacting with several representations simultaneously we generate relations with the representations of which we can then interact and repeat this process recursively, thus remaining in a domain of interactions always larger than that of the representation" (*AC*, p. 14). Reflexivity is thus fundamental to Maturana's account not only because the autopoietic operations of a unity specify for it a world but also because the system's reflexive doubling back on its own representations generates the human subject as an observer.

What about consciousness? Maturana seldom uses this word, preferring to talk instead about "thinking" and "self-consciousness." Thinking occurs in a state-determined nervous system when neurophysiological processes can interact "with some of its own internal states as if these were independent entities." This recursive circling "corresponds to what we call thinking" (*AC*, p. 29). To get from "thinking" to "self-consciousness" requires language, according to Maturana. In the same way that perception does not consist of information from the environment passing into the organism, so language does not consist of someone giving information to someone else. Rather, when an observer uses language, this acts as a trigger for the observer's interlocutor, allowing the interlocutor to establish an orientation within his or her domain of interactions similar to the orientation of the speaker. Only when two entities have largely overlapping domains—for example, when they are both humans sharing similar cultures and beliefs—is it possible for them to achieve the illusion that communication between them has occurred.

From this description, it is apparent that Maturana explains language by simply extending to the linguistic realm the same ideas and terminology he uses to explain perception—an explanation that, in my view, fails to account for some of the distinctive features of language. Shortly we will have an opportunity to look critically at this view of language. For the moment, it permits us to understand Maturana's view of self-consciousness. Self-consciousness arises when the observer "through orienting [linguistic] behavior can orient himself towards himself, and then generate communicative *descriptions* that orient him toward his *description* of this self-orientation." The observer generates self-consciousness, then, when he endlessly

describes himself describing himself. "Thus discourse through communicative *description* originates the apparent paradox of self-description: *self-consciousness,* a new domain of interactions" (*AC,* p. 29). Because Maturana understands self-consciousness solely in linguistic terms, seeing it as an emergent phenomenon that arises from autopoietic processes when they recursively interact with themselves, consciousness for him becomes a epiphenomenon rather than a defining characteristic of the human as an autopoietic entity. The activity of cerebration represents only a fraction of the total autopoietic processes, and self-consciousness represents only a fraction of cerebration. Thus the theory implicitly assigns to consciousness a much more peripheral role than it does to autonomy and individualism. In this respect, autopoietic theory points toward the posthuman even as it reinscribes the autonomy and individuality of the liberal subject.

The complex relation of autopoietic theory to liberal humanism becomes even more apparent when we ask how the theory attempts to establish a foundational ground for itself. As we saw in chapter 1, liberal humanism (in C. B. Macpherson's reading of it) grounds itself on the notion of possessive individualism, the idea that subjects are individuals first and foremost because they own themselves. The equivalent foundational premise in autopoietic theory is the idea that living systems are living because they instantiate organizational closure. It is precisely this closure that guarantees the subject will operate as an autonomous individual. But how is it that Maturana (or anyone else) knows that organizational closure exists? Is the claim that autopoietic closure is *intrinsically* a feature of living systems, or is it how a human observer *perceives* living systems, including itself? This question lies coiled around the brainstem of autopoietic theory, layered into its evolutionary history through its founding distinctions between qualities intrinsic to autopoietic processes and qualities attributed to them by an observer. If the theory says that the observer creates the system by drawing distinctions, it risks undercutting the ontological primacy of organizational closure. If it says that autopoietic processes are an essential feature of reality, it risks undercutting its epistemological radicalism. Faced with this Scylla and Charybdis, Maturana at first steered toward relativism and then, as its dangers loomed closer, changed course and steered toward the absolutism of autopoietic processes existing in reality as such.

So in "Biology of Cognition," the earlier essay in *Autopoiesis and Cognition,* Maturana often wrote as if it is the observer's action that distinguishes an autopoietic unity from its background or medium. "Although a distinction performed by an observer is a cognitive distinction and, strictly, the unity thus specified exists in his cognitive domain as a description, *the ob-*

server in his discourse specifies a metadomain of descriptions from the perspective of which he established a reference that allows him to speak as if a unity . . . existed as a separate entity" (*AC*, p. xxii, emphasis added). This implies the autopoietic unity exists as a distinction that is performed by the observer rather than as an entity that could exist in the absence of an observer. However, in "Autopoiesis: The Organization of the Living," the second and later essay in *Autopoiesis and Cognition*, Maturana and Varela wrote as if an autopoietic unity has the ability to constitute itself independent of an observer. Autopoietic machines, through "*their interactions and transformations . . . continuously regenerate and realize the network of processes (relations) that produced them,*" in the process constituting themselves "*as a concrete unity in the space in which they (the components) exist by specifying the topological domain of* [the autopoietic machine's] *realization of such a network*" (*AC*, p. 79). Here the operation of the autopoietic entity itself—rather than a distinction drawn by an observer—creates the space in which the entity exists. Even more explicit is the claim that individuality comes from the processes themselves rather than from the actions of an observer. "Autopoietic machines have individuality; that is, by keeping their organization as an invariant through its continuous production they actively maintain an identity which is independent of their interactions with an observer" (*AC*, p. 80).

It is not surprising that the issue continues to be debated in autopoietic theory, for it admits of no easy solution. In Maturana's desire to found autopoiesis on something more than an observer's distinction, we can see him trying to pull away from the tar baby of his own reflexive language. Relevant for our purposes is not so much the resolution to this dilemma (as if there could be a definitive resolution!) or even the demonstration that the theory's founding moves make it vulnerable to deconstructive critique. Rather, the important point here is that the foundational ground for establishing the subject's autonomy and individuality has shifted from self-possession, with all of its implications for the imbrication of the liberal subject with industrial capitalism. Instead, these privileged attributes are based on organizational closure (the system closes on itself, by itself) or on the reflexivity of a system recursively operating on its own representations (the observer's distinctions close the system). Closure and recursivity, then, play the foundational role in autopoietic theory that self-possession played in classic liberal theory. The emphasis on closure is visually apparent in the computer simulations, called tessellation automata, that Varela created to illustrate autopoietic dynamics. In contrast to the artificial-life programs that will be discussed in chapter 9, the point of tessellation simulations is to find out

how boundaries close on themselves, how they are maintained when interacting with other tessellation automata, and how and when boundaries break down, which in autopoietic theory is equivalent to death. In this description we see the affinity of autopoiesis not for industrial capitalism (which Maturana frequently excoriates) but for utopian anarchy. Autonomy is important not because it serves as the (paradoxical) foundation for market relations but because it establishes a sphere of existence for the individual, a location from which the subject can ideally learn to respect the boundaries that define other autopoietic entities like itself. This emphasis on closure, autonomy, and individuality also changes what count as primary concerns. When the existence of the world is tied to an observer, the urgent questions revolve around how to maintain boundaries intact and still keep connection with a world that robustly continues to exist regardless of what we think about it.

These changes from liberal humanism also bring with them limitations that are distinctively different from those of first-wave cybernetics. Whereas first-wave philosophies tended to obscure the importance of embodiment and the observer, autopoietic theory draws its strength precisely from its emphasis on these attributes. Its Achilles' heel, by contrast, is accounting for living systems' explosive potential for transformation. The very closure that gives autopoietic theory its epistemological muscle also limits the theory, so that it has a difficult time accounting for dynamic interactions that are not circular in their effects. A prime example, in my view, is the convoluted and problematic way that Maturana treats language. Consistent with his emphasis on circularity, he prefers to talk not about language but "languaging," a process whereby observers, acting solely within their own domain of interactions, provide the triggers that help other observers similarly orient themselves within their domains. Autopoietic theory sees this exchange as a coupling between two independent entities, each of which is formed only by its own ongoing autopoietic processes. As this description shows, the theory is constantly in danger of solipsism, a danger it both acknowledges and attempts to avoid by protesting that it is *not* solipsistic. The main reason the theory adduces for not being solipsistic is its acknowledgment of "structural coupling," the phrase used to denote an organism's interaction with the environment. Even if we grant that this move rescues the theory from solipsism, the theory still seriously understates the transformative effects that language has on human subjects. We have only to recall the term that Maturana employs for a language-using subject—"the observer"—to see how curiously inert and self-enclosing is his view of language.

What drops from view in Maturana's account is the *active* nature of linguistic interactions. Researchers from Jean Piaget on have shown that a child's neural hardware continues to develop after birth in conjunction with the linguistic and social environment in which the child is embedded. In light of this work, it is misleading to talk about the process of active shaping through language simply as an entity "orienting" itself with the aid of an environmental trigger. To appreciate just how active this process is, we can look at instances where it has been short-circuited and the child thus consequently fails to develop normally. In *Mindblindness: An Essay on Autism and Theory of Mind,* Simon Baron-Cohen argues this is what happens with autistic children.[20] Somehow the shaping mechanisms fail to direct neural development, and as a result the child is unable to create an internal scenario that would explain why others act as they do. For such children, Baron-Cohen argues, the world of social interactions is chaotic and unpredictable because they suffer from "mindblindness," an inability to imagine for others the emotions and feelings they themselves have. Autopoietic theory, in its zeal to construct an autonomous sphere of action for self-enclosing entities, formulates a description that ironically describes autistic individuals more accurately than it does normally responsive people. For the autistic person, the environment is indeed merely a trigger for processes that close on themselves and leave the world outside.

In the next section, we will turn to a discussion of how autopoietic theory treats evolution. Like language, evolution represents another area where Maturana's version of autopoietic theory fails to come to terms with the dynamic, transformative nature of the interactions between living systems and their environments. From there we will explore the split that develops between Maturana and Varela. While Maturana continued to replicate his original formulation of the theory, Varela and others became increasingly interested in changing the theory so that it could better account for dynamic interactions. Keeping many of the central insights of autopoietic theory, Varela added new material and reworked some assumptions in the seriated pattern of innovation and replication we have seen at work in other sites. One effect of these changes was to allow elements of autopoietic theory to be integrated into contemporary cognitive science and especially artificial life, which will be the focus of my discussion of third-wave cybernetics in chapter 9.

At this point, a summary may be useful of how autopoietic theory contributes to our evolving stories of (1) the reification of information, (2) the construction of the cyborg, and (3) the transformation of the human into the posthuman. First, whereas first-wave cybernetics played a large role

in divesting information of its body, autopoietic theory draws attention to the fact that "information," so defined, is an abstraction that has no basis in the physically embodied processes constituting all living entities. Autopoiesis thus swerves from the trajectory traced in chapters 3 and 4 with regard to information, insisting that information without a body does not exist other than as an inference drawn by an observer. Second, whereas first-wave cybernetics envisioned the cyborg mostly as an amalgam between the organic and the mechanical, autopoietic theory uses its expanded definition of life to speculate on whether social systems are alive. The paradigmatic cyborg for autopoiesis is the state, not the kind of mechanical human imagined by Bernard Wolfe or Philip K. Dick. Third, autopoietic theory preserves the autonomy and individuality characteristic of liberal humanism, but it sees thinking as a secondary effect that arises when an autopoietic entity interacts with its own representations. Self-consciousness, a subset of thinking, is relegated to a purely linguistic effect. The grounding assumptions for individuality shift from self-possession to organizational closure and the reflexivity of a system recursively operating on its own representations.

A status report, then: information's body is still contested, the empire of the cyborg is still expanding, and the liberal subject, although more than ever an autonomous individual, is literally losing its mind as the seat of identity.

Autopoiesis and Evolution

It is no accident that evolution is a sore spot for autopoietic theory, for the theory was designed to correct what Maturana and Varela saw as an overemphasis on evolution and reproduction as the defining characteristics of life. Over and over, they argue that evolution and reproduction are logically and practically subordinate to autopoiesis. "Reproduction and evolution are not essential for living organisms," they assert in *Autopoiesis and Cognition* (*AC*, p. 11). They are even more opposed to defining living organisms in terms of genetic code. As Varela made clear in a retrospective assessment, he and Maturana were consciously aware of wanting to provide an alternative account of life, an account that would *not* depend in any important way on the idea of a genetic code. "The notion of autopoiesis was proposed . . . with the intention of redressing what seemed to us to be a fundamental imbalance in the understanding of living organization." In correcting this imbalance, they had two interrelated goals. Along with creating a theory of the living that would debunk the current emphasis on DNA as

the "master molecule" of life, they also wanted to insist on the holistic nature of living systems.[21]

Varela is willing to admit that perhaps they erred on the side of overemphasizing autopoiesis at the expense of genetics. By contrast, Maturana became if anything more confirmed in his opposition as time went on. Many critics, including Richard Lewontin, Evelyn Fox Keller, Lily Kay, Richard Doyle, and others, have commented on the distortions created in modern biology by the present overemphasis on DNA.[22] But few go as far as Maturana. In the 1980 article "Autopoiesis: Reproduction, Heredity, and Evolution," a recapitulation of autopoietic theory, he wrote, "I claim that nucleic acids do not determine hereditary and genetic phenomena in living systems, and that they are involved in them, like all other cellular components, according to the particular manner in which they integrate the structure of the living cell and participate in the realization of its autopoiesis."[23] Let us grant that modern biology overemphasizes the role of DNA and that DNA is, as Maturana points out in this passage, only one of many cellular components involved in reproduction. Does he nevertheless go too far in the other direction by insisting that everything be subordinated to autopoiesis?

The problems created by subordinating everything to autopoiesis can be seen in *The Tree of Knowledge*, an account of autopoiesis written for a general audience.[24] As the opening diagram indicates, Maturana and Varela envision each chapter leading into the next, with the final one coming back to the beginning, so that the form of the book recapitulates the circularity of autopoietic theory. "We shall follow a rigorous conceptual itinerary," they announce in the introduction, "wherein every concept builds on preceding ones, until the whole is an indissociable network" (p. 9). In *Autopoiesis and Cognition*, Maturana commented that he and Varela were unable to agree on how to contextualize the theory, so he wrote the introduction by himself. Now, seven years later, Varela is less his student and more an accomplished figure in his own right. This is the last work the two men will coauthor together; Varela has already begun to head in a different direction. The divergences in their viewpoints are accommodated through a clever visual device. Certain key ideas are separated from the text and put into boxes. Each box has a cartoon figure representing the speaker. Maturana's figure wears heavy glasses and is noticeably older than Varela's, so it is easy to identify which is which. Sometimes Maturana's figure authorizes the boxed comments, sometimes Varela's, and sometimes both together. Even without the boxes, it is not difficult to discern that Varela's voice is stronger in *The Tree of Knowledge* than in *Autopoiesis and Cognition*.

I take Varela's Buddhist orientation to be the inspiration behind what the authors announce as a central idea, "all doing is knowing and all knowing is doing" (p. 27). They illustrate the concept by constructing the book as a circle, starting their discussion with unicellular organisms (first-order systems), progressing to multicellular organisms with nervous systems (second-order systems), and finally coming to cognitively aware humans who interact through language (third-order systems). Pointing out that humans in turn are composed of cells, they close the circle by nesting first- and second-order systems within third-order systems, thus joining the doing of autopoiesis with the knowing of cognitively aware creatures. Autopoiesis is the governing idea connecting systems at all levels, from the single cell to the most complex thinking being. "What defines [living systems] is their autopoietic organization, and it is in this autopoietic organization that they become real and specify themselves at the same time" (p. 48). Traversing this path, the "doing" of the reader—the linear turning of pages during the reading—is to become a kind of "knowing" as the reader experiences the organization characteristic of autopoiesis through a textual structure that circles back on itself.

The problem comes when the authors try to articulate this circular structure together with evolutionary lineages. In evolution, lineage carries the sense both of continuity (traced far enough back, all life originates in single-cell organisms) and of qualitative change (different lines branch off from one another and follow separate evolutionary pathways). Whereas in autopoiesis the emphasis falls on circular interactions, in evolution lines proliferate into more lines as speciation takes place through such mechanisms as genetic diversity and differential rates of reproductive success. The tension between evolutionary lines of descent and autopoietic circularity becomes apparent in the authors' claim that autopoiesis is conserved at every point as organisms evolve. To describe the changes taking place, the authors use the term "natural drift." There seems to be a natural drift in "natural drift," however, and in later passages "natural drift" becomes "structural drift." If structure changes, what does it mean to say that autopoiesis is conserved? Here they fall back on the structure/organization distinction that they had previously used in *Autopoiesis and Cognition*. "*Organization* denotes those relations that must exist among the components of a system for it to be a member of a specific class. *Structure* denotes the components and relations that actually constitute a particular unity and make its organization real" (p. 47). Interestingly, they use a mechanical rather than a biological analogy to illustrate the distinction. A toilet's parts can be made of wood or plastic; these different materials correspond to dif-

ferences in structure. Regardless of the material used, however, the toilet will still be a toilet if it has a toilet's organization. The analogy is strangely inappropriate for biology. All life is based on the same four nucleotides; hence for living organisms, it is not the material that changes but the way the material is organized.

What does it mean, then, to say that autopoiesis is conserved? According to the authors, it means that organization is conserved. And what is organization? Organization is "those relations that must exist among the components of a system for it to be a member of a specific class" (p. 47). These definitions force one to choose between two horns of a dilemma. Consider the case of an amoeba and a human. Either an amoeba and a human have the same organization, which would make them members of the same class, in which case evolutionary lineages disappear because all living systems have the same organization; or else an amoeba and a human have different organizations, in which case organization—and hence autopoiesis—must not have been conserved somewhere (or in many places) along the line. The dilemma reveals the tension between the conservative circularity of autopoiesis and the linear thrust of evolution. Either organization is conserved and evolutionary change is effaced, or organization changes and autopoiesis is effaced.

The strain of trying to articulate autopoiesis with evolution is perhaps most apparent in what is not said. Molecular biology is scarcely mentioned and then only in contexts that underplay its importance—a choice consistent with Maturana's claim that heredity does not depend on nucleic acids. There is an additional problem in bringing up molecular biology, for any discussion of DNA coding would immediately reveal that the distinction between structure and organization cannot be absolute—and if this distinction goes, autopoiesis is no longer conserved in evolutionary processes. For if organization is construed to mean the biological classes characterized as species, then it is apparent that organization changes as speciation takes place. If organization means something other than species, then organization ceases to distinguish between different kinds of species and simply becomes the property of any living system. Conserving organization means conserving life, a fact that may be adequate for autopoiesis to qualify as a property of living systems but does nothing to articulate autopoiesis with evolutionary change.

The essential problem here is not primarily one of definitions, although the problem becomes manifest at these sites in the text because definitions are used to anchor the argument, which otherwise drifts off into such nebulous terms as "natural drift." Rather, the difficulties arise because of Mat-

urana's passionate desire to have something conserved in the midst of continuous change. Leaving aside the problems with his explanation of structure and organization, that something is basically the integrity of a self-contained, self-perpetuating system that is operationally closed to its environment. In Maturana's metaphysics, the system closes on itself and leaves historical contingency on the outside. Even when he is concerned with the linear branching structures of evolution, he turns this linearity into a circle and tries to invest it with a sense of inevitability. Seen as a textual technology, *The Tree of Knowledge* is an engine of knowledge production that vaporizes contingency by continuously circulating it within the space of its interlocking assumptions.[25]

Nowhere is the divergence of Varela and Maturana since 1980 clearer than on this point. While Varela moved on to other issues and ways of thinking about them, Maturana continued to occupy essentially the same position and to use the same language as in *Autopoiesis and Cognition.* Clearly Maturana has a more intense and long-lasting commitment to the original formulation of autopoiesis than does Varela. Not coincidentally, Maturana regards himself as the father of the theory, whereas he sees Varela's role as more tangential. In a 1991 article titled "The Origin of the Theory of Autopoietic Systems," he claims credit as the creator of the theory and says that Varela was very much a collaborator who appeared on the scene after the basic ideas had been formulated. "Strictly, Francisco Varela did not contribute to the development of the notion of autopoiesis," Maturana wrote. "This notion was developed between 1960 and 1968. Francisco was my student as an undergraduate during the years 1966 and 1967 in Santiago, then he went to Harvard where he was from 1968 to 1970, when he returned to Chile to work with me in my laboratory in the Faculty of Sciences in Santiago." Although Varela's *Principles of Biological Autonomy* clearly shows that Varela did most of the actual computer work in creating tessellation automata, Maturana claims credit for this idea too. He wrote, "During the year 1972, I proposed one day to make a computer program that would generate an autopoietic system in a graphic space as the result of generating in that space certain elements like molecules."[26] In *Principles,* Varela acknowledged that Maturana was among those "who have influenced this book so pervasively" that their thought was woven into it throughout, but he also wrote in "Describing the Logic of the Living," his 1981 retrospective assessment of autopoiesis, that "the notion of autopoiesis was proposed by Humberto Maturana *and myself.*"[27] This jostling for position, especially when a theory has proven to be historically important, is of course common in almost every field, and particularly in scientific communities, where

great emphasis is placed on being the first to discover something. I mention it here not in any way to diminish the contributions of either Maturana or Varela but to contextualize the fact that Varela moved on to other ways of thinking about autopoiesis while Maturana continued to write in much the same vein as when he had started.

The Voice of the Other: Varela and Embodiment

After *The Tree of Knowledge,* Varela increasingly moved away from the closure that remains a distinctive feature of autopoiesis. The change can be seen in "Describing the Logic of the Living: The Adequacy and Limitations of the Idea of Autopoiesis," his contribution to the important 1981 collection edited by Milan Zeleny: *Autopoiesis: A Theory of Living Organization.* While stressing that he continues to see autopoiesis as very valuable because it "pointed to a neglecting of *autonomy* as basic to the living individual," Varela also criticizes autopoiesis for going both too far and not far enough (p. 37). It went too far, in his view, in becoming a paradigm not just for biological organisms but for social systems as well. Insisting that autopoiesis should not be confused with organizational closure in general, he points out that "the definition of autopoiesis has some precision because it is based on the idea of *production* of components, and this notion of production cannot be stretched indefinitely without losing all of its power" (pp. 37–38). Although cells and animals clearly do physically produce the components that instantiate their organization, social systems do not. Departing from Maturana on this point, Varela would restrict autopoiesis to where, in his view, it is most applicable, to the "domain of cells and animals" (p. 38).

Autopoiesis did not go far enough in building a bridge between its approach and the first-wave emphasis on information flow, teleology, and behavior. "We did not take our criticism far enough to *recover* a non-naive and useful role of information notions in the descriptions of living phenomena," he wrote. Conceding that information, coding, and messages can be "valid explanatory terms," he suggests that they might serve as complementary modes of description for autopoiesis (p. 39). Although he continues to maintain that autopoiesis is logically *necessary* to a complete explanation, it may not be "sufficient to give a satisfactory explanation of living *phenomena* on both logical *and* cognitive grounds" (p. 44). "There was, evidently, a need in [*Autopoiesis and Cognition*] to overemphasize a neglected side of a polarity" (p. 39). To posit an analogous situation in literature, imagine trying to explain how to read a Shakespearean sonnet by starting out with a de-

scription of cellular processes. Logically, it is true that the behavior resulting in reading the sonnet has to originate in cellular processes, but one does not need to be a literature teacher to see that a "chunked," higher-level description would be much more useful.

What Varela argues for, finally, is a dual system of explanation. The operational explanation would emphasize the physical concreteness of actual processes; the symbolic or systems-theoretic explanation would emphasize more abstract ideas that help to construct the system at a higher level of generality. Even so, this "duality of explanation" should "remain in full view" as an antidote to those in computer science and systems engineering who mistake a symbolic description for an operational one, for example by considering that "information and information processing are in the same category as matter and energy." In this respect Varela remains fiercely loyal to autopoiesis. "To the extent that the engineering field is prescriptive by design, this kind of epistemological blunder is still workable. However, it becomes unbearable and useless when exported from the domain of prescription to that of description of natural systems. . . . To assume in these fields that information is some *thing* that is transmitted, that symbols are *things* that can be taken at face value, or that purposes and goals are made clear by the systems themselves is all, it seems to me, nonsense. . . . Information, *sensu strictu,* does not exist. Nor do, by the way, the laws of nature" (p. 45).

In more recent work, Varela and his coauthors provide a positive dimension to this critique of disembodied information. They explore the *constructive* role of embodiment in ways that go importantly beyond autopoiesis. Although autopoietic theory implicitly privileges embodiment in its emphasis on actual biological processes, it has little to say about embodied action as a dynamic force in an organism's development. It is precisely this point that is richly elaborated by Varela and his coauthors in their concept of "enaction."[28] Enaction sees the active engagement of an organism with the environment as the cornerstone of the organism's development. The difference in emphasis between enaction and autopoiesis can be seen in how the two theories understand perception. Autopoietic theory sees perception as the system's response to a triggering event in the surrounding medium. Enaction, by contrast, emphasizes that perception is constituted through perceptually guided actions, so that movement within an environment is crucial to an organism's development. As Varela further explained in "Making It Concrete: Before, During, and After Breakdowns," enaction concurs with autopoiesis in insisting that perception must not be understood through the viewpoint of a "pre-given, perceiver-

independent world." Whereas autopoietic theory emphasizes the closure of circular processes, however, enaction sees the organism's active engagement with its surroundings as more open-ended and transformative. A similar difference informs the views of cognition in the two theories. For autopoiesis, cognition emerges from the recursive operation of a system representing to itself its own representations. Enaction, by contrast, sees cognitive structures emerging from "recurrent sensory-motor patterns."[29] Hence, instead of emphasizing the circularity of autopoietic processes, enaction emphasizes the links of the nervous system with the sensory surfaces and motor abilities that connect the organism to the environment.

Embedded in the idea of enaction is also another story about what consciousness means, a story different from that articulated by autopoietic theory. In *The Embodied Mind: Cognitive Science and Human Experience,* Varela and his coauthors take the Buddhist-inspired point of view that the "self" is a story consciousness tells itself to block out the fear and panic that would ensue if human beings realized there is no essential self. Opposed to the false unity and self-presence of grasping consciousness is true awareness, which is based on actualizing within the mind an embodied realization of the person's ongoing processes. We saw that autopoietic theory invokes the "domain of the observer" as a way to integrate common-sense perceptions with the theory's epistemological radicalism, a move that ended up deconstructing the liberal humanist subject in some respects but recuperating it in others. By contrast, in enaction, consciousness is seen as a cognitive balloon that must be burst if humans are to recognize the true nature of their being. The thrust of *The Embodied Mind* is to show that cognitive science has already been headed in this direction and to interpret the significance of this trajectory in the framework of Buddhist philosophies of emptiness and the not-self. Here the boundaries of the liberal subject are not so much penetrated, stretched, or dissolved as they are revealed to have been an illusion all along. In contrast to the anxiety and nostalgia that Wiener and Maturana expressed when confronted with the loss of the liberal subject, Varela, speaking in a voice now not conjoined with his teacher and mentor, celebrates the moment when the self drops away and awareness expands into a realization of its true nature. No longer Wiener's island of life in a sea of entropy or Maturana's autonomous circularity, awareness realizes itself as a part of a larger whole—unbounded, empty, and serene.

What marks this realization as something other than a mystical vision is Varela's insistence that the most advanced research in Western cognitive science points toward the same conclusion. Referencing such works as R. Jackendoff's *Consciousness and the Computational Mind* and Marvin

Minsky's *Society of Mind* (about which we will hear more in chapter 9), he and his coauthors show that contemporary models of cognition implicitly deconstruct the notion of a unified self by demonstrating that cognition can be modeled through discrete and semiautonomous agents. Each agent runs a modular program designed to accomplish a specific activity, operating relatively independent of the others. Only when conflicts occur between agents does an adjudicating program kick in to resolve the problem. In this model, consciousness emerges as an epiphenomenon whose role it is to tell a coherent story about what is happening, even though this story may have little to do with what is happening processurally. These models posit the mind, Varela wrote, "not as a unified, homogenous unity, nor even as a collection of entities, but rather as *a disunified, heterogeneous, collection of processes*" (p. 100).

In "Making It Concrete," Varela expands this line of thought by showing how Minsky's "society of mind" model can be combined with nonlinear dynamics to give an account of living systems in action. He continues to insist on the importance of the concrete and embodied. "The concrete is not a step towards anything: it is how we arrive and where we stay." Reminiscent of the autopoietic theory's claim that processes happen always and only in the present, he remarks that "it is in the immediate present that the concrete actually lives" (p. 98). To show how Minsky's model is incomplete, he points out that "it is not a model of neural networks or societies; it is a model of the cognitive architecture that abstracts (again!) from neurological detail and hence from the web of the living and of lived experience." "What is missing here," he continues, "is the detailed link between such agents and the incarnated coupling, by sensing and acting, which is essential to living cognition" (p. 99).

The question he poses is how the mind can move smoothly from one agent processing its program to another agent running quite another program. To answer this question satisfactorily, he suggests, we need to link these abstractions with embodied processes. He proposes a "readiness to action" that in effect constitutes a microidentity. As an example, he imagines a man walking down the street, and Varela sketches the kind of behavior associated with this microidentity. Suddenly the man realizes he has left his billfold behind in the last store he visited. Instantly a different microidentity kicks in, geared toward a search operation rather than a leisurely stroll down the street. How does one get from the microidentity of "stroll" to the microidentity of "intense search"? The answer, Varela speculates, involves chaotic, fast dynamics that allows emergent self-organizing structures to arise. In linking the dynamics of self-organizing structures with

microidentities, Varela is following a line of thought vigorously pursued by Zeleny and others, who want to join autopoietic theory with the dynamics of self-organizing systems.[30] The idea is to supplement autopoietic theory so that it can also more adequately account for change and transformation and also to specify the mechanisms and dynamics through which an autopoietic system progresses from one instant in the present to another. These revisions aim to jog autopoietic theory out of its relentless repetitive circularity by envisioning a living organism as a fast, responsive, flexible, and self-organizing system capable of constantly reinventing itself, sometimes in new and surprising ways. In this turn toward the new and unexpected, autopoietic theory begins to look less like the homeostasis of the first wave and more like the self-evolving programs that will be discussed in chapter 9 as exemplars of third-wave cybernetics.

As autopoietic theory continues to evolve, what are likely to be the enduring contributions of autopoiesis as Maturana originally formulated it? In my view, these will certainly include the following: his emphasis on the concreteness and specificity of embodied processes; his insistence that the observer must be taken into account, with all the implications this has for scientific objectivism; his distinction between allopoietic and autopoietic systems, and the ethical implications bound up with making this distinction; and his insight that, in a literal sense, we make a world for ourselves by living it.

In one of his more radical moments, Maturana used the insights of autopoiesis to push toward a formulation that, taken out of context, sounds solipsistic indeed: "We do not see what we do not see, and what we do not see does not exist."[31] In context, he is always careful to qualify this apparent solipsism by pointing out that a world outside the domain of one observer may exist for others, as when I see a large stationary object that a frog cannot perceive. In this way, the world's existence is recuperated in a modified sense—not as an objectively existing reality but as a domain that is constantly enlarging as self-conscious (scientific) observers operate recursively on their representations to generate new representations and realizations. If this isn't exactly the "scientific quest for new knowledge," it nevertheless allows for a qualified vision of scientific progress.

But what if "the observer" ceases to be constructed as a generic marker and becomes invested with a specific psychology, including highly idiosyncratic and possibly psychotic tendencies? Will the domains of self-conscious observers fail to stabilize external reality? Will the uncertainties then go beyond questions of epistemology and become questions of ontology? Will the observation that "what we do not see does not exist" sink deep

into the structure of reality, undermining not only our ability to know but the ability of the world to be? To entertain these suppositions is to enter into the world as it is constructed in the literary imagination of Philip K. Dick. Writing contemporaneously with Maturana but apparently with no knowledge of autopoietic theory, Dick is obsessed with many of the same issues. In turning from Maturana's radical epistemology to Dick's radical ontology, we will follow our evolving stories of the reification of information, the construction of the cyborg, and the emergence of the posthuman into a phantasmagoric territory that continues to exist only as long as an observer thinks it does. And what observers Dick's characters turn out to be!