Chapter 4

Cognition as a biological phenomenon

The previous chapter presented the primary basis for our theoretical orientation, but our own understanding initially developed through a different path. The rationalistic orientation of our prior training in science and technology made the foundations of hermeneutics and phenomenology nearly inaccessible to us. Before we could become open to their relevance and importance we needed to take a preliminary step towards unconcealing the tradition in which we lived, recognizing that it was in fact open to serious question.

For us, this first step came through the work of Humberto Maturana, a biologist who has been concerned with understanding how biological processes can give rise to the phenomena of cognition and language. Beginning with a study of the neurophysiology of vision, which led to the classic work on the functional organization of the frog's retina,[1] he went on to develop a theory of the organization of living systems[2] and of language and cognition.[3]

---

[1]Maturana et al., "Anatomy and physiology of vision in the frog" (1960).

[2]Maturana, "The organization of the living: A theory of the living organization" (1975); Maturana and Varela, *Autopoiesis and Cognition* (1980).

[3]See, for example, Maturana, "Neurophysiology of cognition" (1970), "Cognitive strategies" (1974), "Biology of language: The epistemology of reality" (1978). Much of Maturana's theory was developed in conjunction with Francisco Varela, whose own work is further developed in Varela, *Principles of Biological Autonomy* (1979), "Living ways of sense making: A middle way approach to neurosciences" (1984), and *El Arbol de Conocimiento* (forthcoming). Maturana was deeply influenced by work in cybernetics by von Foerster and others in the Biological Computer Laboratory at the University of Illinois. The book *Cybernetics of Cybernetics* (von Foerster, 1974),

In comparing Maturana's orientation to Heidegger's, it is important to recognize that they began in very different traditions. As we mentioned in the previous chapter, there is a long history within philosophy of viewing mental and physical descriptions as applying in incommensurate domains. In approaches based on this 'dualism' it is taken for granted that mental predications (such as "X knows that Y" or "X perceives a Y") are not expressible in terms of physical descriptions of a nervous system. Having made this assumption, it becomes a confusion of levels to ask whether a particular physical activity of the nervous system is a 'perception' or whether a certain state is one in which the organism 'knows' some 'fact.'

Among the scientists who work in areas such as neurophysiology and artificial intelligence, however, it is a strongly held working hypothesis that there is a systematic and recurrent relationship between the two domains. It is assumed that "X sees a red spot" can be correlated with a particular pattern of activity in the retina and visual cortex, or that "John believes that Brutus killed Caesar" can be associated with a particular pattern of data in John's brain, viewed as a computer with appropriate software and storage devices. Few researchers adopt the naive approach of looking for immediate correlations between the mental and the physical except in peripheral functions like the image manipulation done by the retina. Usually the argument is based on an analogy to computer programs, in which the organization of the software provides a level of 'functional description' that is abstracted away from the specifics of the physical implementation. An entity counts as being explained when its behavior can be described in terms of a compositional analysis that postulates parts that are functionally identified—that play functionally defined roles in its operation.

It is possible to adopt the position, as Chomsky at times does,[4] that theories of cognition can deal purely with 'competence,' characterizing the behavior of the cognitive system while making no hypothesis concerning the generation of that behavior by mechanisms. Most cognitive scientists, however, find this stance too restrictive. As scientists, they take it for granted that all observable phenomena are ultimately explainable in terms of mechanistic systems operating according to regular laws. A cognitive theory needs to deal with the causal principles by which these systems operate, not just with abstract characterizations of the behavior they generate. The appropriate level of description for the causal regularities may be an abstract level of software, rather than the physical descriptions of individual components, but nevertheless it has the goal of explaining *why*

---

produced in that laboratory, provides a broad insight into its work. Since we have been most directly influenced by Maturana's writings we will refer primarily to them and to him.

[4]See, for example, Chomsky, *Reflections on Language* (1975).

things happen, not just characterizing what happens.

In creating such an explanation, it has generally been taken for granted that the distinctions and relations that are applied in describing the mental domain will form a basis for examining the structures in the domain of causal mechanism. Maturana, beginning as an experimental neurophysiologist, came to realize that this naive formulation was inadequate and that it obscured the phenomena he wanted to study. However, he also wanted to adhere to the scientific tradition of explanation in terms of deterministic physical systems, asking how such systems might give rise to cognitive phenomena.[5] Much of his writing, therefore, is directed toward revealing the pervasiveness of the pre-understanding that biologists and cognitive scientists bring to bear, and toward opening possibilities for a different understanding.

In his writings, Maturana introduces a good deal of new terminology which seems puzzling and difficult on first reading. He does this quite consciously, because he recognizes that the old terminology carries within it a pre-understanding that is a trap for new understanding. For example, in describing their use of the word 'autopoiesis' Maturana and Varela say:

> Curiously, but not surprisingly, the introduction of this word proved of great value. It simplified enormously the task of talking about the organization of the living without falling into the always gaping trap of not saying anything new because the language does not permit it. We could not escape being immersed in a tradition, but with an adequate language we could orient ourselves differently and, perhaps, from the new perspective generate a new tradition. — Maturana and Varela, *Autopoiesis and Cognition* (1980), p. xvii.

We introduce much of Maturana's terminology, without attempting to give definitions (indeed our own theory of language denies the possibility of giving precise definitions). The network of meanings will gradually evolve as the different ideas are developed and the links of their interdependence laid out. We cannot in these few pages give a complete or balanced account of Maturana's work. We have chosen instead to emphasize those aspects that were most critical in the development of our own understanding, trusting that for many readers this introduction will motivate the challenging but rewarding task of reading the original sources.

---

[5] As will become obvious in this chapter and throughout the book, the words 'cognitive' and 'cognition' are used in quite different ways by different writers. We will not attempt to give a simple definition of Maturana's use, but will clarify it through the discussion in the chapter.

4.1 The closure of the nervous system

In neurophysiological studies of vision, the traditional assumption (based on a rationalistic philosophy of cognition) was that the activity in the optic nerve was a direct representation of the pattern of light on the retina. The work of Maturana, Lettvin, McCulloch, and Pitts on "Anatomy and physiology of vision in the frog" (1960) challenged this, demonstrating that over large areas of the retina to which single fibers of the optic nerve were connected, it was not the light intensity itself but rather the pattern of local variation of intensity that excited the fiber. There was, for example, one type of fiber that responded best to a small dark spot surrounded by light. When triggered, it led to activity appropriate for catching a fly in the location corresponding to the spot. It became apparent that at least some of the cognitive processes that we would interpret as relevant to the survival of the frog actually take place within its visual system, not deeper in its neuroanatomy.

In trying to extend this research to color vision, Maturana, Uribe, and Frenk ("A biological theory of relativistic color coding in the primate retina," 1968) made observations that led to further questioning the relation between perception and the world being perceived. Making use of a simple observation that had been noted for many years, they argued that theories associating colors directly with wavelengths on the spectrum were inadequate. When a stick is illuminated by a white light from one side and a red light from the other, it casts two shadows, one of which appears red (against a generally pink background) and the other of which appears *green*. If we ask about the objective 'thing' being observed, there is no light with a spectrum of wavelengths normally called green; only various shades of red, white, and pink. However, Maturana and other researchers have postulated that the patterns of neural activity produced are the same as those produced by light of a single wavelength normally called green.[6] The presence of 'green' for the nervous system is not a simple correlate of the presence of certain wavelengths of light, but the result of a complex pattern of relative activity among different neurons.

This example was one of many that led Maturana to question the validity of our commonsense understanding of perception. On the naive view, there is some objectively recognizable property of a thing in the environment, and our perception of it is a capturing of that property in our mind. This idea is contained in the very words used in description:

> When Jerry Y. Lettvin and I wrote our several articles on frog vision..., we did it with the implicit assumption that we were

[6]The original work in this area is described in Land, "The retinex theory of color vision" (1977).

> 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), and the animal could use the information obtained in its perception to compute a behavior adequate to the perceived situation. This assumption of ours appeared clearly in our language. We described the various kinds of retinal ganglion cells as feature detectors, and we spoke about the detection of prey and enemy. — Maturana, "Biology of cognition" (1970), p. xii.

Further examination of visual phenomena, however, suggested a different orientation. In order to deal with the seemingly fundamental perceptual category of color it was necessary to give explanations in terms of relative patterns of activity within the nervous system.

> I soon realized in my research that my central purpose in the study of color vision could not be the study of a mapping of a colorful world on the nervous system, but rather that it had to be the understanding of the participation of the retina (or nervous system) in the generation of the color space of the observer. — Maturana, "Biology of cognition" (1970), p. xii.

Perception, in other words, must be studied from the inside rather than the outside—looking at the properties of the nervous system as a generator of phenomena, rather than as a filter on the mapping of reality.

Maturana describes the nervous system as a closed network of interacting neurons such that any change in the state of relative activity of a collection of neurons leads to a change in the state of relative activity of other or the same collection of neurons. From this standpoint, the nervous system does not have 'inputs' and 'outputs.' It can be *perturbed* by structural changes in the network itself, and this will affect its activity, but the sequence of states of the system is generated by relations of neuronal activity, as determined by its structure.

When light strikes the retina, it alters the structure of the nervous system by triggering chemical changes in the neurons. This changed structure will lead to patterns of activity different from those that would have been generated without the change, but it is a misleading simplification to view this change as a perception of the light. If we inject an irritant into a nerve, it triggers a change in the patterns of activity, but one which we would hesitate to call a 'perception' of the irritant. Maturana argues that all activity of the nervous system is best understood in this way. The focus should be on the interactions within the system as a whole, not on the structure of perturbations. The perturbations do not determine what happens in the nervous system, but merely trigger changes of state. It is the structure of the perturbed system that determines, or better, *specifies* what structural configurations of the medium[7] can perturb it.

From this perspective, there is no difference between perception and hallucination. If the injected irritant creates a pattern of neural activity identical to that which would be produced by heat applied to the area served by the nerve, then there is no neurophysiological sense to the question of whether the heat was really 'perceived' or was a 'hallucination.' At first, this refusal to distinguish reality from hallucination may seem far-fetched, but if we think back to color vision it is more plausible. The question of whether the shadow in the stick experiment was 'really green' is meaningless once we give up the notion that the perception of green corresponds in a simple way to a pattern of physical stimuli. In giving a scientific explanation of the operation of the nervous system at the physical level, we need to explain how the structure of the system at any moment generates the pattern of activity. The physical means by which that structure is changed by interaction within the physical medium lie outside the domain of the nervous system itself.

Of course an observer of the nervous system within its medium can make statements about the nature of the perturbation and its effect on patterns of activity. For this observer it makes sense to distinguish the situation of an injected irritant from one of heat. But *from the standpoint of the nervous system* it is not a relevant, or even possible, distinction.

Along with this new understanding of perception, Maturana argues against what he calls the 'fallacy of instructive interaction.' 'Instructive interaction' is his term for the commonsense belief that in our interactions with our environment we acquire a direct representation of it—that properties of the medium are mapped onto (specify the states of) structures in the nervous system. He argues that because our interaction is always through the activity of the entire nervous system, the changes are not in the nature of a mapping. They are the results of patterns of activity which, although triggered by changes in the physical medium, are not representations of it. The correspondences between the structural changes and the pattern of events that caused them are historical, not structural. They cannot be explained as a kind of reference relation between neural structures and an external world.

The structure of the organism at any moment determines a *domain of perturbations*—a space of possible effects the medium could have on the sequence of structural states that it could follow. The medium selects among these patterns, but does not generate the set of possibilities. In

[7] Here and throughout this chapter we use the term 'medium' rather than 'environment' to refer to the space in which an organism exists. This is to avoid the connotation that there is a separation between an entity and its 'environment.' An entity exists as part of a medium, not as a separate object inside it.

understanding an organism as a structure-determined system we view it in terms of its components and the interactions among them. The appropriate domain of description is not the behavior of the organism as a unity, but the interlocking behaviors of its physical components.

4.2 Autopoiesis, evolution, and learning

Maturana's understanding of an organism's relation to its environment leads to an epistemological problem. In our culture's commonsense theories of knowledge, what we know is a representation of the external world. Based on information gathered through perception, our brain somehow stores facts, uses them to draw conclusions, and updates them on the basis of experience.

If we look at the nervous system as closed, we must ask how an organism comes to have any knowledge of the world. How can a history of independent perturbations lead to the phenomena of cognition that our introspective intuitions will not let us deny? Maturana seeks to explain the origins of all phenomena of cognition in terms of the phylogeny (species history) and ontogeny (individual history) of living systems. To do this, he must first give an adequate account of the organization of the living. Maturana and Varela characterize the organization of the living as 'autopoietic.'[8] An autopoietic system is defined as:

> ...a network of processes of production (transformation and destruction) of components that produces the components that: (i) through their interactions and transformations continuously regenerate the network of processes (relations) that produced them; and (ii) constitute it (the machine) as a concrete unity in the space in which they (the components) exist by specifying the topological domain of its realization as such a network. — Maturana and Varela, *Autopoiesis and Cognition* (1980), p. 79.

The terms used in this definition, such as 'unity,' 'component,' and 'space,' have technical meanings in Maturana's work. We will not attempt to define them here, but will explicate their meaning in the course of describing the relevance of this definition.

The phenomenon of autopoiesis is quite general. It can apply to systems existing in any domain in which we can identify unities and components. An autopoietic system holds constant its organization and defines its boundaries through the continuous production of its components. If

[8]For a collection of papers by Maturana, Varela, and others on autopoiesis, see Zeleny, *Autopoiesis, a Theory of the Living Organization* (1978).

the autopoiesis is interrupted, the system's organization—its identity as a particular kind of unity—is lost, and the system disintegrates (dies). An autopoietic system that exists in physical space is a living system.[9]

At first sight, this definition may seem irrelevant or trivial. But it is in fact a carefully crafted statement expanding on a simple idea: the essential characteristic of a living system is that it is a collection of components constituting a unity that can live or die. Maturana's central observation is that exactly this simple property leads to the complex phenomena of life. The functioning of an organism as a structure-determined system with the potential of disintegration leads to adaptation and evolution.

The mechanism by which an organism comes to function adequately in its medium is one of selection, which includes both the selection of structural changes within an individual and the selection of individuals by the possibilities of survival and disintegration. A plastic, structure-determined system (i.e., one whose structure can change over time while its identity remains) that is autopoietic will by necessity evolve in such a way that its activities are properly coupled to its medium. Its structure must change so that it generates appropriate changes of state triggered by specific perturbing changes in its medium; otherwise it will disintegrate.

> Learning is not a process of accumulation of representations of the environment; it is a continuous process of transformation of behavior through continuous change in the capacity of the nervous system to synthesize it. Recall does not depend on the indefinite retention of a structural invariant that represents an entity (an idea, image, or symbol), but on the functional ability of the system to create, when certain recurrent conditions are given, a behavior that satisfies the recurrent demands or that the observer would class as a reenacting of a previous one. — Maturana, "Biology of cognition" (1970), p. 45.

Structural coupling is the basis not only for changes in an individual during its lifetime (learning) but also for changes carried through reproduction (evolution). In fact, all structural change can be viewed as ontogenetic (occurring in the life of an individual). A genetic mutation is a structural change to the parent which has no direct effect on its state of autopoiesis until it plays a role in the development of an offspring.

> If... the observer wants to discriminate between learned and instinctive behavior, he or she will discover that in their actual

[9] In later work, Maturana and Varela distinguish autopoiesis, as a property of cellular systems, from a more general property of *operational closure* that applies to a broader class of systems. We will not pursue the distinction here, but it is explicated in Varela's *El Arbol de Conocimiento* (forthcoming).

> realization, both modes of behavior are equally determined in the present by the structures of the nervous system and organism, and that, in this respect, they are indeed indistinguishable. The distinction between learned and instinctive behaviors lies exclusively in the history of establishment of the structures responsible for them. — Maturana, "Biology of language" (1978), p. 45.

The structural coupling generated by the demands of autopoiesis plays the role that we naively attribute to having a representation of the world. The frog with optic fibers responding to small moving dark spots does not have a *representation* of flies. As a result of structural coupling, the structure of the nervous system generates patterns of activity that are triggered by specific perturbations and that contribute to the continued autopoiesis of the frog. Of course, the changes of structure that led to the frog's nervous system would not have been supportive of autopoiesis if the frog had to compete for food in a flyless environment. But it is an error to assume that the structure reflects a knowledge of the existence of flies. An explanation of why a frog catches a fly can be in two different domains. In the domain of the frog as a physical system we can explain how the structure determines the action. In the cognitive domain (discussed below), we can explain how the history of perturbations of the frog (and its ancestors) led to the structure that determines it.

4.3 The cognitive domain

Maturana's primary concern is to understand what 'cognition' can mean when we relate it to the fundamental nature of living systems. Rejecting the metaphor of information processing as the basis for cognition, he replaces the question "How does the organism obtain information about its environment?" with "How does it happen that the organism has the structure that permits it to operate adequately in the medium in which it exists?" An answer to this question cannot be generated by comprehending how a nervous system operates, but must be grounded in a deeper understanding of how cognitive activity is common to all life, and is determined by the underlying phenomenon of autopoiesis. As Maturana observes ("Neurophysiology of cognition," 1970, p. 8), "Living systems are cognitive systems, and living, as a process, is a process of cognition. This statement is valid for all organisms, with and without a nervous system."

What does it mean, then, to understand an organism as a cognitive system?

> A cognitive system is a system whose organization defines a domain of interactions in which it can act with relevance to the maintenance of itself, and the process of cognition is the actual (inductive) acting or behaving in this domain. — Maturana, "Biology of cognition" (1970), p. 13.

A cognitive explanation is one that deals with the *relevance* of action to the maintenance of autopoiesis. It operates in a *phenomenal domain* (domain of phenomena) that is distinct from the domain of mechanistic structure-determined behavior:

> ... as a result of the structural coupling that takes place along such a history, history becomes embodied both in the structure of the living system and the structure of the medium even though both systems necessarily, as structure determined systems, always operate in the present through locally determined processes.... History is necessary to explain how a given system or phenomenon came to be, but it does not participate in the explanation of the operation of the system or phenomenon in the present. — Maturana, "Biology of language" (1978), p. 39.

As observers we can generate descriptions of the activity of living systems in either of two non-intersecting domains. One description deals with the structure of the system and how that structure determines behavior. Such a description is essentially ahistorical. It does not matter how the system came to be that way, only that it is. We can at the same time describe (as observers of a history of changes within the structure and the medium) the pattern of interactions by which the structure came to be, and the relationship of those changes to effective action. It is this second domain of explanation that Maturana calls 'cognitive.' The cognitive domain deals with the relevance of the changing structure of the system to behavior that is effective for its survival.

It is therefore in this cognitive domain that we can make distinctions based on words such as 'intention,' 'knowledge,' and 'learning.' As mentioned above, philosophers have been careful to distinguish 'mental predicates' such as these from the physical predicates that apply to the organisms or machines embodying the phenomena they describe. For Maturana the cognitive domain is not simply a different (mental) level for providing a mechanistic description of the functioning of an organism. It is a domain for characterizing effective action through time. It is essentially temporal and historical. A statement like "The animal knows X" is not a statement about its state, but a statement about a pattern of actions (past and projected). One of Maturana's main objectives is to overcome the tendency (imposed on us by our language) to treat mental terms as though they could meaningfully be taken as descriptions of state or structure.

In making a distinction between the domain of cognition and the domain of the nervous system as a structure-determined system, Maturana clarifies an issue that has been at the heart of behaviorist approaches to cognition. The behaviorist describes the behavior of the organism (its responses) as a function of the sequence of states of the environment—the relevant stimuli, reinforcements, and punishments. In the most extreme form of behaviorism, it is assumed that stimuli and responses can be described externally without reference to the structure of the organism, and that all of the interesting recurrences of behavior can be explained by the patterning of the events. There are many who would call themselves behaviorists (or 'neobehaviorists'[10]) who postulate an internal state of an organism as well as the record of inputs and outputs. What is common to behaviorist approaches is the focus on the organism as a behaving entity (a unity, in Maturana's terms) subject to external stimuli, rather than as a composite to be understood through the interactions of its components.

Although he recognizes the significance of a domain that is independent of the structure of the cognitive system, Maturana rejects the behaviorist view, arguing that we cannot deal with 'organism' and 'environment' as two interacting independent things. We cannot identify stimuli that exist independently of the unity and talk about its history of responses to them. The unity itself specifies the space in which it exists, and in observing it we must use distinctions within that space.

4.4 Consensual domains

The sources of perturbation for an organism include other organisms of the same and different kinds. In the interaction between them, each organism undergoes a process of structural coupling due to the perturbations generated by the others. This mutual process can lead to interlocked patterns of behavior that form a *consensual domain.*

> When two or more organisms interact recursively as structurally plastic systems, ... the result is mutual ontogenic structural coupling.... For an observer, the domain of interactions specified through such ontogenic structural coupling appears as a network of sequences of mutually triggering interlocked conducts.... The various conducts or behaviors involved are both arbitrary and contextual. The behaviors are arbitrary because they can have any form as long as they operate as

[10]For example, see Suppes, "From behaviorism to neobehaviorism" (1975).

> triggering perturbations in the interactions; they are contextual because their participation in the interlocked interactions of the domain is defined only with respect to the interactions that constitute the domain.... I shall call the domain of interlocked conducts...a *consensual domain.*— Maturana, "Biology of language" (1978), p. 47.

For example, when the male and female of a species develop a sequence of mutual actions of approach and recognition in a mating ritual, we as observers can understand it as a coherent pattern that includes both animals. Our description is not a description of what the male and female (viewed as mechanisms made up of physical components) do, but a description of the mating dance as a pattern of mutual interactions. The generation of a consensual domain is determined by the history of states and interactions among the participants (and their progenitors) within the physical domain. However, as observers of this behavior we can distinguish a new domain in which the system of behaviors exists. The consensual domain is reducible neither to the physical domain (the structures of the organisms that participate in it) nor to the domain of interactions (the history by which it came to be), but is generated in their interplay through structural coupling as determined by the demands of autopoiesis for each participant.

Maturana refers to behavior in a consensual domain as 'linguistic behavior.' Indeed, human language is a clear example of a consensual domain, and the properties of being arbitrary and contextual have at times been taken as its defining features. But Maturana extends the term 'linguistic' to include *any* mutually generated domain of interactions. Language acts, like any other acts of an organism, can be described in the domain of structure and in the domain of cognition as well. But their existence *as language* is in the consensual domain generated by mutual interaction. A language exists among a community of individuals, and is continually regenerated through their linguistic activity and the structural coupling generated by that activity.

Language, as a consensual domain, is a patterning of 'mutual orienting behavior,' not a collection of mechanisms in a 'language user' or a 'semantic' coupling between linguistic behavior and non-linguistic perturbations experienced by the organisms.

Maturana points out that language is connotative and not denotative, and that its function is to orient the orientee within his or her cognitive domain, and not to point to independent entities. An observer will at times see a correspondence between the language observed and the entities observed, just as there is a correspondence between the frog's visual system and the existence of flies. But if we try to understand language purely within the cognitive domain, we blind ourselves to its role as orienting behavior, and make inappropriate assumptions about the necessity and nature of reference.

> The basic function of language as a system of orienting behavior is not the transmission of information or the description of an independent universe about which we can talk, but the creation of a consensual domain of behavior between linguistically interacting systems through the development of a cooperative domain of interactions.— Maturana, "Biology of language," p. 50.

The role of 'listening' in generating the meaning of an utterance is closely related to Maturana's explanation of consensual domains, and will be discussed at length in Chapter 5.

4.5 The observer and description

At this point, it is useful to take a step back and apply Maturana's orientation to what we ourselves are doing. In the act of writing down these words on paper we are engaged in linguistic activity. In using language, we are not transmitting information or describing an external universe, but are creating a cooperative domain of interactions. Our own use of language carries a possibly misleading pre-understanding.

There is a naive view (as discussed in Chapters 2 and 5) that takes language as conveying information about an objective reality. Words and sentences refer to things whose existence is independent of the act of speaking. But we ourselves are biological beings, and the thrust of Maturana's argument is that we therefore can never have knowledge about external reality. We can have a structure that reflects our history of interactions in a medium, but that medium is not composed of 'things' that are knowable. We can talk about a world, but in doing so we act as 'observers':

> An observer is a human being, a person, a living system who can make distinctions and specify that which he or she distinguishes as a unity... and is able to operate as if he or she were external to (distinct from) the circumstances in which the observer finds himself or herself. Everything said is said by an observer to another observer, who can be himself or herself.— Maturana, "Biology of language," p. 31.

As observers, we generate *distinctions* in a consensual domain. A description in any domain (whether it be the domain of goals and intention, or that of physical systems) is inevitably a statement made by an observer to another observer, and is grounded not in an external reality but in the consensual domain shared by those observers. Properties of things (in fact the recognition of distinct things at all) exist only as operational distinctions in a domain of distinctions specified by an observer. When we talk about systems and their medium, components, and structural change, we speak as if there were external things and properties. This is an inescapable result of using language, but it is always a speaking 'as if,' not an ontological claim.

This idea that all cognitive distinctions are generated by an observer (and are relative to the nature of that observer) is not new to Maturana. Köhler, for example, in his classical book *Gestalt Psychology* (1929), argued that phenomena—i.e., the way the world presents itself to the naive observer in everyday situations—are not objective or subjective by virtue of whether they arise from internal or external events, since all knowable events are in one sense internal (resulting from internal experiences and neurophysiological events). More recent work in systems theory and cybernetics[11] also challenges the naive acceptance of modes of interpretation that assume the objectivity of observation. What is different and crucial in Maturana's discourse is the recognition that distinctions lie in a consensual domain—that they presuppose some kind of social interaction in which the observer is engaged:

> The linguistic domain as a domain of orienting behavior requires at least two interacting organisms with comparable domains of interactions, so that a cooperative system of consensual interactions may be developed in which the emerging conduct of the two organisms is relevant for both.... The central feature of human existence is its occurrence in a linguistic cognitive domain. This domain is constitutively social.
> — Maturana, "Biology of cognition" (1970), pp. 41, xxiv.

In denying the possibility of subject-independent objective knowledge, Maturana does not adopt the solipsistic position that our discourse can deal ultimately only with our subjective thoughts and feelings. By virtue of being a discourse it lies in a consensual domain—a domain that exists for a social community. Reality is not objective, but neither is it individual:

> ...cultural differences do not represent different modes of treating the same objective reality, but legitimately different cognitive domains. Culturally different men live in different cognitive realities that are recursively specified through their living in them.... The question of solipsism arises only as a

[11] See, for example, Pask, *Conversation Theory* (1976) and *Conversation, Cognition and Learning* (1975), and von Foerster, *Cybernetics of Cybernetics* (1974).

> pseudo-problem, or does not arise at all, because the necessary condition for our possibility of talking about it is our having a language that is a consensual system of interactions in a subject dependent cognitive domain, and this condition constitutes the negation of solipsism.— Maturana, "Cognitive strategies" (1974), p. 464.

In going on to deal with consciousness, Maturana again emphasizes its continuity with other phenomena of cognition, rather than seeing it as a fundamentally different capacity. He sees consciousness as generated through the operation of the consensual domain in which language is generated. Language (shared, not private) is prior to conscious thought.

4.6 Domains of explanation

The relevance of Maturana's work to the design of computers lies in his account of how biological organisms function. It was a critical perturbation to our understanding of computers because it provided a domain of concrete examples of systems that were *mechanistic* but not *programmed.* To understand the importance of this, it is useful to look at an example.

Examining a newborn baby's ability to get food, we see a remarkable collection of extremely successful behaviors. A cry gets mother's attention, the 'rooting' reflex (a light pressure on one cheek produces a head turn to that side) positions the baby's mouth over a nipple, and the sucking actions express milk. If an AI scientist set out to build such a program, he or she might propose that the baby be provided with a set of 'goals,' such as 'drink milk' and 'get nipple in mouth,' a set of 'operators' such as 'cry,' 'turn head,' and 'suck,' and a model of the world that sets them into appropriate correspondence. A more sophisticated model might even include a model of mother's goals and plans, so that the cry could be analyzed as an attempt to evoke an appropriate plan of action on her part.

But of course all of this is irrelevant to the actual mechanisms at work. The baby, like every organism, has a complex set of reflexes whose purposes can be explained in terms like those above, but whose functioning does not depend on representations, planning, or analysis. The result is behavior that is successful for a particular coupling with the medium, but is limited in its range. If the particular actions don't work, there is no generalized ability to come up with 'other ways to eat.' In Maturana's terms, we can describe behavior in either the cognitive domain (in which purposes and coupling are central) or the domain of the mechanism as a structure-determined system (in which the actual reflex paths are the key).

In trying to build computer systems, it has often been argued that we must move beyond 'baby' systems, which simply do what is appropriate because their structure is properly coupled. If the computer can manipulate an explicit model of the goals and potential actions, then it can infer possible action sequences that were not initially programmed but that lead to the desired goals. Greater flexibility should come from specifying the goals and operators instead of the course of action.

To some extent this approach is valid, but in another way it is shortsighted. It assumes that the programmer (or 'knowledge engineer') can articulate an explicit account of the system's coupling with the world—what it is intended to do, and what the consequences of its activities will be. This can be done for idealized 'toy' systems and for those with clearly circumscribed formal purposes (for example programs that calculate mathematical formulas). But the enterprise breaks down when we turn to something like a word processor, a time-sharing system, or for that matter any system with which people interact directly. No simple set of goals and operators can delimit what can and will be done. We might, for example, note that 'goals' of the word processor include allowing a person to compare two pieces of text, to rapidly skim through a document, to copy fragments from one document to another, to move quickly to some part of a document, etc. These might be satisfied by 'window' and 'scrolling' mechanisms, but they will not operate by explicit analysis of the user's goals. The person selects among basic mechanisms that the machine provides, to get the work done. If the mechanisms don't do what is needed, others may have to be added. They will often be used in ways that were not anticipated in their design.

Similarly, the effects of different 'operators' cannot be fully described or anticipated when they elicit responses by people (or even by other computer systems). We can define the domain of perturbations (the space of possible effects the interaction can have on the system), but we cannot model how the system's activity will engender them.

The most successful designs are not those that try to fully model the domain in which they operate, but those that are 'in alignment' with the fundamental structure of that domain, and that allow for modification and evolution to generate new structural coupling. As observers (and programmers), we want to understand to the best of our ability just what the relevant domain of action is. This understanding guides our design and selection of structural changes, but need not (and in fact cannot) be embodied in the form of the mechanism.

In Chapters 8 and 12 we will explore more fully the consequences of Maturana's approach for the design of computer programs, and for the discourse about machines and intelligence.