

2 Horizons for the Enactive Mind: Values, Social Interaction, and Play

Ezequiel A. Di Paolo, Marieke Rohde, and Hanne De Jaegher

2.1 Introduction

Almost two decades since the publication of *The Embodied Mind* (Varela, Thompson, and Rosch 1991), the term *enactive* has moved out of relative obscurity to become a fashionable banner in many regions of cognitive science. It has found its way into diverse areas, from education and human-computer interaction, to autonomous robotics and consciousness studies. On the surface, this acceptance indicates the success of the ideas articulated by Varela and his colleagues, and their view of the mind with its emphasis on the role of embodied experience, autonomy and the relation of co-determination between cognitive agents and their world. Theirs was not only an achieved synthesis of existing criticisms to a predominantly computationalist paradigm, but also the articulation of a set of postulates to move these ideas forward. Indeed, the increasing use of enactive terminology serves as an indication that the time is ripe for a new era in cognitive science. To a great extent, we believe this to be so.

However, on closer inspection, a significant variety of meaning is revealed in the use of the word “enactive” (as happens with closely associated terms such as *autonomous*, *embodied*, *situated*, and *dynamical*). The label sometimes indicates only the partial adoption of enactive views, vaguely connected to the ideas in *The Embodied Mind*. In the worst cases, we see the raising of implausible hybrids risking self-contradiction in their mixture of the new and the old. There seems to be a lack of consensus about what constitutes enactivism or embodied cognitive science in general (Wilson 2002). Enactive has often been taken simply as synonymous of active, embodied as synonymous of physical, dynamical as synonymous of changing, and situated as synonymous of exchanging information with the environment, all properties that could be claimed by practically every robot, cognitive model or theory proposed since

symbolic artificial intelligence (AI) first made its debut as the theoretical core of cognitive science about fifty years ago. This situation can lead to confusion and eventually to the loss of meaning attached to these terms—indeed, a perceived ambiguity between revolution and reform was already noticed by early commentators (Dennett, 1993).

We find at least two reasons for this situation, both indicating pressing problems that must be addressed if enactive cognitive science is to get off the ground. The first one is a watering down of the original ideas of enactivism by their partial adoption or sublimation into other frameworks. The second, related reason is a genuine lack of enactive proposals to advance open questions in cognitive science that motivate more traditional frameworks, such as the problems of higher-level cognition. These reasons lead to the misappropriation of the previously mentioned keywords through the acceptance of the lessons of enactivism, but only for a restricted range of influence. In the opinion of many, the usefulness of enactive ideas is confined to the “lower levels” of human cognition. This is the “reform-not-revolution” interpretation. For instance, embodied and situated engagement with the environment may be sufficient to describe insect navigation, but it will not tell us how we can plan a trip from Brighton to La Rochelle. Or enactive-like ideas could well account for complex skills such as mastering sensorimotor contingencies in visual perception (O’Regan and Noë 2001), or becoming an expert car driver (Dreyfus 2002), but—important though these skills are—they remain cognitively marginal (Clark and Toribio 1994) and fall short of explaining performances such as preparing for a mathematics final or designing a house. For some researchers, enactive ideas are useful but confined to the understanding of sensorimotor engagements. As soon as anything more complex is needed, we must somehow recover newly clothed versions of representationalism and computationalism (Clark and Toribio 1994; Clark 1997; Clark and Grush 1999; Grush 2004).

We would do wrong in ignoring such positions. They are good indicators of what is at the core of the struggle between traditional and unorthodox temperaments in cognitive science today. Indeed, the current situation serves as a reminder of the dangerous fate that fresh and radical ideas may suffer: that of dilution into a background essentially indistinguishable from that which they initially intended to reject. We believe that it is mistaken to conclude that what enactivism cannot yet account for must necessarily be explained using an updated version of old ideas with a debatable success record. But it will remain tempting to do so *as long as the principal tenets and implications of enactivism remain insufficiently clear*. It would also be

wrong to ignore arguments that show the limitations of enactivism. These challenges reveal how much is left to be done. *Enactivism is a framework that must be coherently developed and extended.*

For this reason, in trying to answer the question “What is enactivism?” it is important not to straightjacket concepts that may still be partly in development. Some gaps may not yet be satisfactorily closed; some contradictions may or may not be only apparent. We should resist the temptation to decree solutions to these problems simply because we are dealing with definitional matters. The usefulness of a research program also lies with its capability to grow and improve itself. It can do so only if problems and contradictions are brought to the center and we let them do their work. For this, it is important to be engendering rather than conclusive, to indicate horizons rather than boundaries.

There are still many important areas in enactive cognitive science that demand serious development. These remain the stronghold of traditional conceptions. Most of the underdeveloped areas within the enactive approach involve higher levels of cognitive performance: thinking, imagining, interpreting the behavior of others, and so on. For as long as enactive ideas are taken as filling in details or as playing a contextual role in the explanation of such phenomena, the situation will not change.

We dedicate this chapter to clarifying the central tenets of enactivism and exploring some of the themes currently under development. In this exercise, following the logic of the central ideas of enactivism can sometimes lead to unexpected hypotheses and implications. We must not underestimate the value of a new framework in allowing us to *formulate questions in a different vocabulary*, even if satisfactory answers are not yet forthcoming. Implicitly, the exploration of these questions and possible answers is at the same time a demonstration of the variety of methods available to enactivism, from phenomenology, to theory/experiment cycles, and to the synthesis of minimal models and validation by construction—an additional thread that runs through this chapter and that we will pick up again in the discussion.

In particular, after introducing the five core ideas of enactivism, we focus on value generation and question the coherence of the idea of a *value system* in cognitive architectures (both computationalist and embodied) and similar modular structures whose function is to generate or judge the *meaning* of a situation. This question allows us to highlight right from the start one of the main differences between enactive and traditional views: a grounding of notions such as values and meaning. Many influential theories in cognitive science make use of the idea that value or meaning

is some information appraised by an internal module within an agent's cognitive architecture, whereas in an enactive perspective, meaning is inseparable from the whole of context-dependent, life-motivated, embodied activity, without being at all a hazy concept beyond the reach of scientific understanding. We also explore, continuing on the issue of the origins of meaning, the field of social cognition, the focus of many recent phenomenologically inspired criticisms (Thompson 2001; Gallagher 2001, 2005). Our exploration leads us toward a middle way between individualistic and holistic views of social interaction and to highlighting the central role played by the temporality of social engagements in generating and transforming social understanding at different time scales through joint participation. In the final part, we take a speculative look at the embodied capability to manipulate the meaning of concrete situations by exploring the role of play in the development of human cognition. These explorations do not attempt to be complete, nor do they put the whole of human cognition within the reach of enactivism and forever banish representational/computational explanations. But they do extend the conceptual horizon and allow us to formulate the problem of higher cognitive performance in an alternative, enactive way.

2.2 The Core of Enactivism

It would be misleading to think of the enactive approach as a set of all radically novel ideas. It is much rather a synthesis of some new but also several old themes that mutually support each other. Overall, enactivism may be construed as a kind of nonreductive, nonfunctionalist naturalism. It sees the properties of life and mind as forming part of a continuum and consequently advocates a scientific program that explores several phases along this dimension.

Among the predecessors to enactivism we find, for example, Piaget's theory of cognitive development through sensorimotor equilibration (Piaget 1936, 1967), Poincaré's theory of the active role of movement in the construction of spatial perception (Poincaré 1907), Goldstein's theory of the self-actualizing organism (Goldstein [1934] 1995), and others. The very term "enactive" has been similarly used before, for example by Bruner in the 1960s, to describe knowledge that is acquired and manifested through action (Bruner 1966). Equally, we find philosophical affinities with existential phenomenology (Heidegger 1962; Merleau-Ponty 1962), with Eastern mindfulness traditions, with Hans Jonas's biophilosophy (Jonas 1966), and with pragmatic thinkers such as Dewey (1929). Current

compatibilities can be also found with many embodied and dynamical systems ideas in contemporary cognitive science (Beer 2000; Chiel and Beer 1997; Thelen and Smith 1994; Hutchins 1995a; Juarrero 1999; Kelso 1995), neuroscience (Bach-y-Rita et al. 1969; Damasio 1994; Skarda and Freeman 1987; Engel, Fries, and Singer 2001), evolutionary biology (Lewontin 1983; Oyama 2000), and AI/robotics (Beer 2003; Brooks 1991; Harvey et al. 1997; Nolfi and Floreano 2000; Winograd and Flores 1986). Some of these connections are made explicit in *The Embodied Mind*, others have been elaborated later in the literature, and still others remain to be better established.

What is the core of the enactive approach? Views that take cognition as embodied and situated, or take experience seriously, or explore the purchase of dynamical systems ideas, will all share something with enactivism. But to call them enactive just because there is some conceptual overlap may only contribute to a meaningless proliferation of the term. This is unless we can show both that (1) such views share or are developed from a basic core of enactive ideas, and (2) extensions to these ideas do not result in irresolvable contradictions with this basic core. We can identify five highly intertwined ideas that constitute the basic enactive approach (Varela, Thompson, and Rosch 1991; Thompson 2005): *autonomy*, *sense-making*, *emergence*, *embodiment*, and *experience*. Partially implying each other, these ideas sit on the blind spots of traditional views. We will not attempt to disentangle all of their connections in order to obtain a set of perfectly independent postulates. Indeed, the internal relations between these concepts speak for the strength of their association under a single banner.

2.2.1 Autonomy

Living organisms are autonomous—they follow laws set up by their own activity. Fundamentally, they can be autonomous only by virtue of their self-generated identity as distinct entities. A system whose identity is fully specified by a designer and cannot, by means of its own actions, regenerate its own constitution, can only follow the laws contained in its design, no matter how plastic, adaptive, or lifelike its performance. In order for a system to generate its own laws, it must be able to build itself *at some level of identity*. If a system “has no say” in defining its own organization, then it is condemned to follow an externally given design like a railroad track. As such, it may be endowed with ways of changing its behavior depending on history, but at some level it will encounter an externally imposed functional (as opposed to physical) limitation to the extent to which it can

change itself. This can be avoided only if the system's limitations result partly from its own processes.

The autonomy (or freedom) of a self-constituted system is by no means unconstrained (being able to influence one's own limitations does not imply being able to fully remove them; on the contrary, it means being able to set up new ways of constraining one's own actions). Hans Jonas (1966) speaks of life as sustaining a relation of *needful freedom* with respect to its environment. Matter and energy are needed to fuel metabolism. In turn, by its constant material turnover, metabolism sustains its form (its identity) by dynamically disassociating itself from specific material configurations.

It should be clear that by expressions like "self-constitution" and "generating its own laws" no mysterious vitalism is intended. However, the acceptance of an operational concept of emergence (discussed shortly) is implied. By saying that a system is self-constituted, we mean that its dynamics generate and sustain an identity. An *identity* is generated whenever a precarious network of dynamical processes becomes operationally closed. A system is operationally closed if, for any given process *P* that forms part of the system (1) we can find among its enabling conditions other processes that make up the system, and (2) we can find other processes in the system that depend on *P*. This means that at some level of description, the conditions that sustain any given process in such a network always include those conditions provided by the operation of the other processes in the network, and that the result of their global activity is an identifiable unity in the same domain or level of description (it does not, of course, mean that the system is isolated from interactions with the environment). Autonomy as operational closure is intended to describe self-generated identities at many possible levels (Varela 1979, 1997; Di Paolo 2009).

Cognitive systems are also autonomous in an interactive sense in terms of their engagement with their environment as agents and not simply as systems coupled to other systems (Moreno and Etxeberria 2005; Di Paolo 2005). As such, they not only respond to external perturbations in the traditional sense of producing the appropriate action for a given situation, they do in fact actively and asymmetrically *regulate* the conditions of their exchange with the environment, and in doing so, enact a world or cognitive domain.

To view cognitive systems as autonomous is therefore to reject the traditional poles of seeing mind as responding to environmental stimuli on the one hand, or as satisfying internal demands on the other—both of

which subordinate the agent to a role of obedience to external or internal factors. It is also to recognize the “ongoingness” of sensorimotor couplings that lead to patterns of perception and action twinned to the point that the distinction is often dissolved. Autonomous agency goes even further than the recognition of ongoing sensorimotor couplings as dynamical and emphasizes the role of the agent in constructing, organizing, maintaining, and regulating those closed sensorimotor loops. In doing so, the cognitive agent plays a role in determining the norms that it will follow, the “game” that is being played.

2.2.2 Sense-Making

Already implied in the notion of interactive autonomy is the realization that organisms cast a web of significance on their world. Regulation of structural coupling with the environment entails a direction that this process is aiming toward: that of the continuity of the self-generated identity or identities that initiate the regulation. This establishes a *perspective on the world* with its own normativity, which is the counterpart of the agent being a center of *activity in the world* (Varela 1997; Weber and Varela 2002; Di Paolo 2005; Thompson 2007). Exchanges with the world are thus inherently significant for the agent, and this is the definitional property of a cognitive system: the creation and appreciation of meaning or *sense-making*, in short.

It will be important to notice already— this issue is treated more extensively in the following section—that sense-making is an inherently active idea. Organisms do not passively receive information from their environments, which they then translate into internal representations. Natural cognitive systems are simply not in the business of accessing their world in order to build accurate pictures of it. They participate in the generation of meaning through their bodies and action often engaging in transformational and not merely informational interactions; *they enact a world*. Enactivism thus differs from other nonrepresentational views such as Gibsonian ecological psychology on this point (Varela, Thompson, and Rosch 1991, 203–204). For the enactivist, sense is not an invariant present in the environment that must be retrieved by direct (or indirect) means. Invariants are instead the outcome of the dialog between the active principle of organisms in action and the dynamics of the environment. The “finding” of meaning must be enacted in a concrete and specific reduction of the dimensions that the organism-environment system affords along the axis of relevance for autonomy; it is always an activity with a *formative* trace, never merely about the innocent extraction of information

as if this was already present to a fully realized (and thus inert) agent. This is another idea that sets the enactive framework apart from more traditional views in cognitive science: a dynamical, biologically grounded, theory of sense-making. Like few notions in the past, this concept strikes at the heart of what is to be cognitive. We will elaborate this point in the next section and show how elusive this way of thinking can be even among researchers who have taken embodiment and situatedness very seriously.

2.2.3 Emergence

The overarching question in cognitive science is: How does it work? For the enactive approach, the connected concepts of autonomy and sense-making already invoke some notion of emergence in addressing this question. Autonomy is not a property of a collection of components, but the consequence of a new identity that arises out of dynamical processes in precarious operational closure. Meaning is not to be found in elements belonging to the environment or in the internal dynamics of the agent, but belongs to the relational domain established between the two.

The idea of emergence has been much debated in various domains from metaphysics to epistemology and has had a furious revival over the last three decades with the advent of the sciences of complexity. Beyond the debates about the possibility of ontological emergence (Kim 1999; Silberstein and McGeever 1999), there is a pragmatic application of the term that stems from the well-understood phenomenon of self-organization. This has served to remove the air of mystery around emergence in order to bring it back in line with a naturalistic project. There is also a demand for emergentist explanations in biology, in which hierarchical organization is all too evident (e.g., genetic regulation, cells, extracellular matrices, tissues, organs, organism, dyads, groups, institutions, societies).

Emergence is used to describe the formation of a novel property or process out of the interaction of different existing processes or events (Thompson 2007; Thompson and Varela 2001). In order to distinguish an emergent process from simply an aggregate of dynamical elements, two things must hold (1) the emergent process must have its own autonomous identity, and (2) the sustaining of this identity and the interaction between the emergent process and its context must lead to constraints and modulation to the operation of the underlying levels.¹ The first property indicates the identifiability of the emergent process whose characteristics are enabled but not fully determined by the properties of the component processes. The second property refers to the mutual constraining between emerging

and enabling levels (sometimes described as circular or downward causation).

We find the clearest example of emergence in life itself. The property of continuous self-production, renewal, and regeneration of a physically bounded network of molecular transformations (autopoiesis) is not to be found at any level below that of the living cell itself. Being a self-sustaining bounded network of chemical transformations is not (it cannot be) the property or the responsibility of single components in this network. The new level is not only autonomous in terms of exhibiting its own identity and laws of transformation; it also introduces, through interaction with its codefined context, modulations to the boundary conditions of the lower-level processes that give rise to it.

This phenomenon repeats itself at various levels in multicellular organisms and in particular animals and humans. Variations on this theme have been used to describe the emergence of the self/nonsel self distinction in immune networks (Stewart and Coutinho 2004); the generation, maintenance, and eventual dissolution of coherent modes of synchronous activity in the brain (Engel, Fries, and Singer 2001; Thompson and Varela 2001); and also between these coherent modes and action/perception cycles (Rodriguez et al. 2001; Le Van Quyen and Petitmengin 2002). Emergent phenomena, as indicated in the previous examples, can be fleeting. Single acts can bear a relation of emergence with respect to their sensorimotor component phases.

Taking emergence seriously makes the enactive approach very skeptical about the localization of function corresponding to one level in specific components at a lower level (homuncularity) and consequently leads to the rejection of “boxology” as a valid method to address the “how does it work” question. Any labeling of subsystemic components and variables with names belonging naturally to properties of emergent levels (e.g., value systems, cognitive maps, emotional modules, mirror neurons) should be treated with extreme caution.

Having said all this, emergence remains problematic, due often to its opaqueness and the ease with which the term can be misused. The weight of explaining how a given phenomenon constitutes a proper case of emergence remains with the supporters of this view. The very blurring of distinctions between levels that the enactive approach criticizes of cognitivism has allowed the latter paradigm to connect personal and subpersonal levels with indiscriminate ease. The properties of higher levels are thus explained in terms of lower-level ones, because they are already magically present there. For the emergentist, instead, the connection and the interaction

between levels becomes a problem to be addressed case by case, often by recourse to complex concepts and tools derived from dynamical systems theory. It is clear that much work is still needed for clarifying and operationalizing the concept of emergence. In this context, synthetic models can prove very valuable as tools for grasping emergent phenomena.

2.2.4 Embodiment

In a concrete and practical sense, a cognitive system is embodied to the extent to which its activity depends nontrivially on the body. However, the widespread use of the term has led in some cases to the loss of the original contrast with computationalism and even to the serious consideration of trivial senses of embodiment as mere physical presence—in this view, a word processor running on a computer would be embodied, (cf. Chrisley 2003). It is easy to miss a fundamental motivation behind embodiment. Nontrivial dependence on the body can easily be construed in functionalist terms, and this falls short of the more radical implications of enactivism. It is not only a question of moving the mind from a highly sheltered realm of computational modules in the head into messy bodily structures. So-called embodied approaches that do not move beyond this first step remain largely functionalist and see the body as yet another information processing device; a convenient way to offload computations that would be too hard to handle by the neural tissue (Clark 1997). This is a Cartesian view of embodiment in its separation between mind as function on the one hand and body as implementation on the other. A similar adopted view is that of the mind as controller and the body as controlled. Despite their tension, these views often go together. By contrast, for the enactivist the body is the ultimate source of significance; embodiment means that mind is inherent in the precarious, active, normative, and worldful process of animation, that the body is not a puppet controlled by the brain but a whole animate system with many autonomous layers of self-constitution, self-coordination, and self-organization and varying degrees of openness to the world that create its sense-making activity.

Indeed, to say that cognition is embodied is to express a tautology—it simply cannot *but* be embodied if we understand the core of cognition as sense-making. The latter goes hand in hand with the conservation of emergent identities (autonomy) ultimately constituted by material processes in precarious conditions (i.e., unable to sustain a ‘function’ independently of each other or indefinitely). In other words, mind is possible because a body is always a decaying body (a fact that cannot be captured in functionalist terms).

For enactivism, therefore, cognition is embodied in a fundamental, non-functional sense although it may still nurture itself by the fascinating examples of how bodily structures and dynamics may be cleverly exploited to resolve complex problems both in human performance (Lenay 2003) and in robots (Pfeifer and Scheier 1999; Salomon 1998). The relevance of the body is not restricted to concrete sensorimotor activities. There is much evidence that higher-level cognitive skills, such as reasoning and problem solving, mental image manipulation, and language use depend crucially on bodily structures (Wilson 2002; Lakoff 1987).

There are enactive accounts of the potential layering of several identities into a more or less integrated body-in-interaction (Varela 1997; Di Paolo 2005, 2009). These can serve to make sense of a further twist to the role played by the body in the case of human cognition—one that could explain the resilience of Cartesian modes of thinking. Even though our bodies are not puppets, to say that we control our bodies is, in a sense, not entirely wrong. We certainly do. But we do so in subtle ways that relate to the emergence of forms of reflexive autonomy, this time of a sociolinguistic nature. Like an alien presence, I set new aims for my body (I decide to embrace the pain of a yoga class, I decide to go on a diet). Being able to support and transform new identities is one way in which the body creates the experience of a self not quite the same as (and sometimes at odds with) the metabolic self. Taken in isolation, this is an experience that nurtures Cartesianism. In fact, the body, by further manipulating its sense-making activity, is capable of putting itself in a novel situation that is partly its own creation. In doing so, it is playing a highly skillful dual role. This is afforded by the plasticity of the human body, but it would not be possible without immersion within a symbolic order and the social mediation that makes our bodies fit to a scheme of control and observation of behavioral and cultural norms thus giving rise to sociolinguistic and narrative selves.

2.2.5 Experience

For enactivism, experience is central both methodologically and thematically. Far from being an epiphenomenon or a puzzle as it is for cognitivism, experience in the enactive approach is intertwined with being alive and immersed in a world of significance. As part of the enactive method, experience goes beyond being data to be explained. It becomes a guiding force in a dialog between phenomenology and science, resulting in an ongoing pragmatic circulation and mutual illumination between the two (Gallagher 1997; van Gelder 1999; Varela 1996, 1999).

Many modern accounts of cognitive activity already take experience seriously. For instance, Dreyfus's defense of nonrepresentational skill acquisition (2002) is based on paying careful attention to the experience of undergoing a process of task improvement. As we make the journey from beginners to experts through practice, not only is skillful performance improved, but experience is also transformed. This is to be expected if embodiment in the enactive sense is taken seriously. If experience and the body-in-interaction were to relate to each other as two mutually external systems, we would expect either an unchangeable or a fleeting relation between our bodies and our experience. Instead we find a lawful relation of bodily and experience transformations. Becoming a wine connoisseur is certainly an achievable goal but expertise in this field (as in any other) is not obtained through gaining the right kind of *information* but through the right kind of *transformation*—one that can only be brought about by appropriate time-extended training (experimenting, making mistakes, and so on). Experience is altered in a lawful manner through the process. It is itself a skillful aspect of embodied activity.

An embodied perspective results in serious attention being paid to isomorphisms between mechanisms and experience. Varela (1999) and van Gelder (1999) provide different, but related, dynamical systems accounts of mechanisms that might underlie the protentive and retentive structure of time consciousness as described by Husserl. Kelly (2000) considers neural models of pointing and grasping that run parallel to Merleau-Ponty's concepts of the intentional arc and maximal grip. Wheeler (2005) explores isomorphic relationships between embodied/embedded accounts of situated action and Heideggerian categories such as the ready-to-hand, breakdowns, and present-at-hand. What is interesting in many of these accounts is that the process of circulation is not one of assimilating scientific hypotheses into phenomenology, but may itself inform phenomenology. This is as it should be in a proper dialog, and such is the methodology advocated by first-person methods in the joint study of experience and brain-body activity (Varela 1996; Lutz 2002).

Experience may also serve the role of clarifying our commitments. Hans Jonas (1966) looks into the world of living beings and sees that life is a process with interiority. Metabolism has all the existential credentials of concerned being. It is precarious, it separates itself from nonbeing, it struggles to keep itself going and preserve its identity, and it relates to the world in value-laden terms. However, the inward aspect of life cannot be demonstrated using our current scientific tools. This does not make it any less factual for Jonas. He knows that all life is connected along evolutionary

chains, and he knows that we ourselves are embodied living creatures with an inner life. This is how we can then know that living beings are forms of existence and that they also have an inner life.

This example is telling, because it already contains a difficult-to-swallow consequence of the dialog between science and experience, which is, at the same time, perhaps its most revolutionary implication. Phenomenologically informed science goes beyond black marks on paper or experimental procedures for measuring data, and dives straight into the realm of personal experience. No amount of rational argument will convince a reader of Jonas's claim that, as an embodied organism, he is concerned with his own existence if the reader cannot see this for himself. Jonas appeals to the performance of a gesture that goes beyond comprehending a scientific text. The implication is that in order to work as a source of knowledge, enactivism will contain an element of personal practice. It is necessary to come back to the phenomenology and confirm that our theories make sense, but this means that sometimes we must become skillful in our phenomenology as well—personally so.

2.3 Values and the Limits of Evolutionary Explanations

The previous section shows that there are certain ideas in cognitive science that the enactive approach clearly rejects, such as homuncularity, boxology, separability between action and perception, and representationalism. In this section, we will revisit some of these themes in a more focused manner.

In everyday life, we experience the world in value-laden terms. This fact is hard to avoid and has been the subject of much philosophical debate. For enactivism, value is simply an aspect of all sense-making, as sense-making is, at its root, the evaluation of the consequences of interaction for the conservation of an identity. Perhaps as a reaction to the subjective overtones of this issue, traditional cognitive science has not dwelled much on the explicit mechanisms involved in value judgment as an inherent aspect of cognitive activity. In general, questions about value or natural purposes have been dealt with separately, preferably with reference to evolutionary history (Millikan 1984): everything living organisms do is ultimately reduced to survival strategies in situations like those encountered by their ancestors, or to the urge to spread their genes as widely as possible. In a more traditional cognitive modeling framework, this idea translates to values being “built-in” by evolution—phylogenetically invariant yardsticks against which actual lifetime encounters are measured and

structured, and from which cognitive mechanisms that are themselves independent of these values deduce the meaning of situations, actions, and perceptions.

Explanations of this kind are in tension with the principles of enactivism, in particular with the concept of sense-making. In this section, we juxtapose such traditional views, in which ultimate ends come in evolutionarily sealed boxes, with an alternative, more enactive view that explains values and meanings as consequences of the kind of dynamical system a living organism is. We discuss an enactive theory of value in its rudimentary form, which is based on the theory of autopoiesis. A number of open questions, such as the explanation of nonmetabolic values or transitions in value-generating mechanisms are raised and implications for computational models of cognition are discussed.

2.3.1 Values: Built-in or Constructed?

Weber and Varela (2002) were the first to suggest a derivation of intrinsic teleology, natural purposes, and the capacity of sense-making from autopoiesis, drawing on Kant's *Critique of Judgment* and Hans Jonas's philosophy of biology (Jonas 1966), and the position argued for here commits to this general idea. In this kind of reasoning, the struggle for continuing autopoiesis—in other words, survival—is at the core of intrinsic teleology and the capacity of sense-making. Even though survival plays a central role in both autopoietic and evolutionary explanations of value (one must first survive in order to reproduce), there are essential differences between the claim that what affects an organism's autopoietic organization is of value and the claim that values are built-in because they benefit survival and hence have been selected for.

If values are built-in, they need to have some form of priority over the living, acting creature, either temporally or logically. Typically, claims about biological traits being built-in are about them being part of the genetic package. "Value" is a term that describes the meaning of organismic behavior, not one of its physiological or mechanistic properties, like, for instance, the blood type. Therefore, the idea of built-in values relies on some kind of a priori *semantics*: parts of the genetic code are thought to execute according to preprogrammed rules and, thereby, generate values. This automated "sense lookup" is not the same as sense-making. Similarly, we are dealing with *pre-factum* evolutionary teleonomy, not with autonomy. Instead of emergence, we find a direct reduction of evaluative function to physical structures. Instead of embodiment, we find abstract principles that are presumed independent of embodied interaction. Finally,

lived experience is subdued as secondary to historical selection pressure—whether value is manifested experientially seems irrelevant. The idea of built-in values and the enactive approach diverge along all those lines.

This may sound like a very black-and-white picture. Maybe not all that living organisms do can be explained through built-in values, but there are surely some basic properties and behaviors that will always benefit survival, such as that oxygen, food, water, and light will always be good for most animals, so what is wrong with claiming that there are some built-in basic values like “water is good,” “light is good,” “this food is good”? The point is not to argue that such norms do not exist across individuals of a species, but rather that they should be searched for on the emergent level of autonomous interaction, not on the level of mechanism. If we imagine that a mechanistic structure inside a living organism were solely responsible for the generation of values, does that mean that the remainder of the organism is value-agnostic, that the values generated by this mechanism are arbitrary? Would that not imply that a mutation of the genetic code that tells the organism that “food is good” could result in the generation of the value “poison is good”? For the mutant system, poison would then be of positive value, just as food was for its ancestor, even if this mutation would eventually kill it, which seems a strange idea. The facts that food and water and light are good and that poison is bad are a result of the kind of system that an organism is here and now and that they are of consequence for its conservation. In this sense, no mutation can create the value “poison is good” without changing the organization of the system so that it thrives on “poison.” The value for this organism would again be “food is good,” not “poison is good.” In other words, “good to eat” in enactive terms is equivalent to “stuff I can turn into more of me.” The organism is an ontological center that imbues interactions with the environment with significance they do not have in its absence, and this significance is not arbitrary. It is dynamically constructed, and that is the essence of the idea of sense-making.

The thrust behind the idea of precoded values, in contrast, is the assumption of an isomorphism between what is genuinely good or bad for the organism and what the executed genetic value programs say is good. Precoded values are thought to predict the effect of lifetime encounters for metabolism, on the basis of phylogenetic history. Therefore, they rely on phylogenetic constancies. However, cases in which we can observe a change of relation between a value and an organism demonstrate the ontological priority of biological autonomy. The most striking examples of such value changes—which can shatter the functionality of established

relations—include illness, perceptual supplementation, and other perturbations to the body (distortions or impairments). Bach-y-Rita et al. (1969) have demonstrated the amazing human capacity to perceive visually, despite a loss of sight, by relaying pixelated images, recorded with a head-mounted camera, to arrays of tactile stimulators. What kind of preexistent, evolutionarily shaped, built-in value mechanism could be responsible for assigning the meaning sighted people make of light patterns to tickling stimuli on the skin when the situation does not correspond to any history of selective pressures?

Or consider a patient who, during the course of a disease, is subjected to increasing dosages of a pharmaceutical agent, with the result that he not only survives dosages of the drug that would be fatal to the average human being, but also that his metabolism relies on the medicine in a way that deprivation would cause his death. The value of this substance for the metabolism is inverted as a consequence of the changes undergone by the organism. But the transformation is not arbitrary. On the contrary, the kind of system that the organism becomes will determine the drug's altered value, and this determination cannot be attributed to a local module, evolutionarily dedicated to the task of assigning meaning, but to the system as a whole. If constancies break down, we observe that local mechanisms gradually undergo a change in how their function relates to meaning such that local processes are not anymore about the same thing they once were when they were selected for. We call this phenomenon *semantic drift*; it comes up again in section 2.3.3.

Even if it is true that specific internal structures play a fundamental role in the value-appraisal process, reducing the latter to the former is a category mistake; it confounds the domains of mechanism and of behavior. To localize the correlated function in these structures is like saying that the speed of a car is in the gas pedal.

2.3.2 Kinds of Values

We propose to define value as *the extent to which a situation affects the viability of a self-sustaining and precarious network of processes that generates an identity*. The most widely discussed and most intensely analyzed such process is autopoiesis, the continuous material regeneration of a self-bounded, self-constructing network of molecular transformations in a far-from-equilibrium situation. Encounters will be good or bad depending on their effect on autopoiesis. Up to this point, our discussion has exclusively argued the case of this basic “metabolic value,” as it seems the least controversial. It now remains to be established what kinds of other processes

might be self-sustaining, precarious, and generate an identity, that is, what other processes might generate values?

Logically, there are two possibilities for value generation by processes other than metabolism itself: value generation alongside autopoiesis and value generation independent of autopoiesis. Both scenarios immediately lead to further questions. If there are self-sustaining precarious processes that generate an identity, but are fully independent of living organisms, where does teleology come from? Can we really say that such processes generate value, and if yes, value for whom or for what? By contrast, if such processes “parasite” on the process of living, how do the values they generate relate to the basic metabolic value? What happens in case of a conflict? The enactive paradigm leaves space for a multitude of possible positions on these matters; these questions are far from settled and this section cannot but present a few existing positions and our own thoughts in progress.

Varela’s own perspective on the organism as a “meshwork of selfless selves” (1991, 1997) can be seen as an exploration into value-generating mechanisms, mainly of the first kind—that is, based on autopoiesis as the most basic form of autonomy and identity generation. Identity generation, for him, entails that an invariant quality is maintained coherently by an operationally closed process whose primary effect is its own sustained production. Varela studied three mechanisms to bring about such processes: autopoiesis (cellular identity), the immune system (multicellular biochemical identity) and the nervous system (neurocognitive identity). He acknowledges the existence or possibility of other levels of identity, reaching from precellular identity (e.g., identity of self-replicating molecules) to sociolinguistic identity and superorganismic identity. Similar ideas are elaborated by Jonas (1966); see also Di Paolo 2009. In a similar spirit, Barandiaran, Di Paolo, and Rohde (2009) have studied the mutual constraints between autonomy (such as a living organism’s metabolic self-production) and agency (embodied sensorimotor behavior).

We want to touch on some examples from a nonexhaustive listing of transitions in value-generating mechanisms (figure 2.1) that we consider particularly important or interesting, drawing on some of Varela’s and Jonas’s ideas. The first three stages of this scale are frequently not treated as distinct. However, it has recently been argued (Di Paolo 2005) that mere autopoiesis, according to the original definition, even though it is sufficient to generate natural teleology and metabolic value, does not entail active appraisal of the corresponding metabolic norms: an autopoietic entity can be robust to perturbations without the logical necessity to

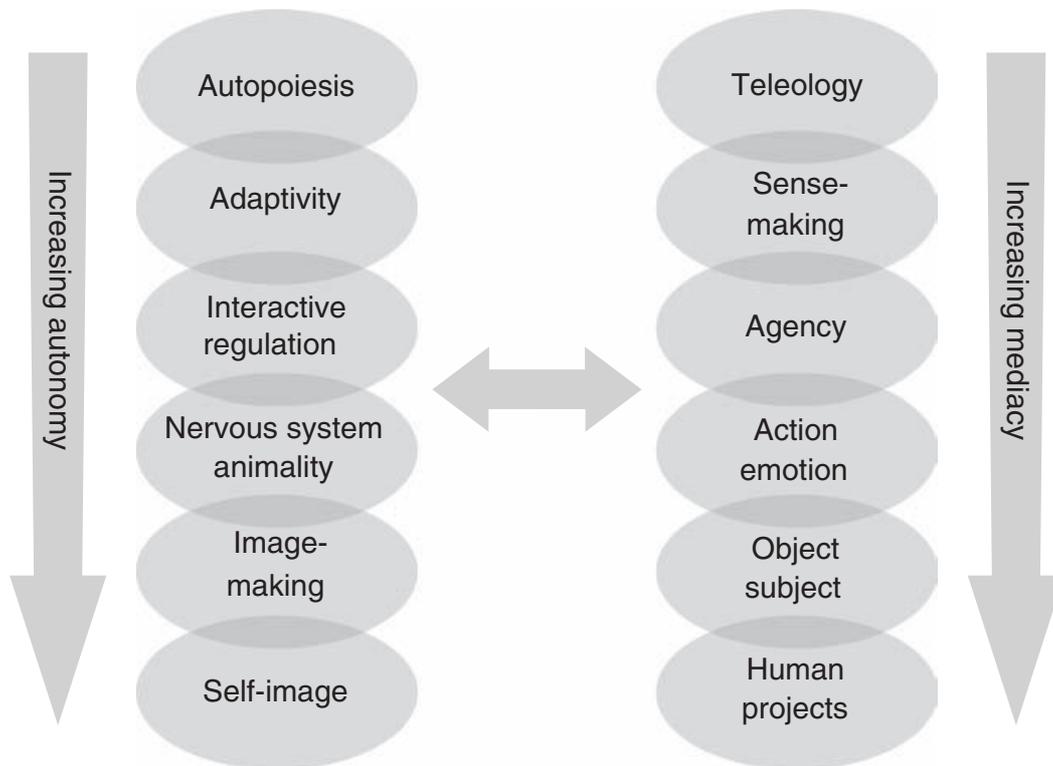


Figure 2.1

Life-mind continuity and the scale of increasing autonomy and mediacy.

actively monitor its own state and act to improve the conditions for continued autopoiesis. Only *adaptive* autopoietic entities that improve the conditions for continued autopoiesis, by actively monitoring their own state, identifying at least some tendencies that bring them closer to the boundary of viability and counteracting these tendencies can be actual “sense-makers.” A similarly subtle distinction is the one between adaptive organisms and interactive regulators (Moreno and Etxeberria 2005): although the former act to counter hostile tendencies by changing their internal organization, the latter act on the environment and thereby exhibit the most fundamental form of agency (cf. Barandiaran, Di Paolo, and Rohde 2009). An example of a just-adaptive organism is the sulfur bacterium that survives anaerobically in marine sediments, whereas bacteria swimming up a sugar gradient would, by virtue of their motion, qualify for minimal agency.

The further stages on the scale (figure 2.1) are largely adopted from Jonas’s work. Animals, through their motility, exhibit the capacity to act and perceive as well as fear or desire for something distal. And humans, through socially mediated capacities such as image-making and ultimately for constructing a self-image, gain the ability to regard situations objec-

tively and define themselves as subjects, to distinguish truth from falsehood, and to experience happiness and frustration (Jonas 1966; Di Paolo 2005). This ladder follows the “gradient of autonomy and mediacy.” It connects increasing degrees of mediation between an urge and its satisfaction to higher degrees of precariousness, and to the consequent liberation of ways to generate values. For instance, only a sense-making organism is capable of mistakes by virtue of the mediacy of urge and satisfaction. A bacterium that swims up the ‘saccharine’ gradient, as it would in a sugar gradient, can be properly said to have assigned significance to a sign that is not immediately related to its metabolism, even though it is still bound to generate meanings solely based on the consequences for its metabolism. With increasing mediacy, the possibilities to create meaning for signs become less and less constrained by the instantaneous metabolic needs of the organism. Such hierarchies of processes bringing about different kinds of identities and values relate to the study of the major transitions in evolution like the evolution of the eukaryote cell, of sex, or of multicellularity (Maynard Smith and Szathmáry 1995). However, even though different organizations of living creatures enable new and more complex kinds of value-generating processes, transitions in structure cannot immediately be equated with transitions in value, the evolution of value-generating processes proceeds in a more gradual and continuous fashion. The exact relation between complication of material organization through processes of reproduction and selection and the evolution of values is largely unexplored territory that certainly deserves future attention.

One of the riddles in this picture is how different kinds of values are tied together to form a unitary self (and whether they always do). By calling the organism a “meshwork of selfless selves,” Varela avoids the answer to this question: “Organism as self, then, cannot be broached as a single process. We are forced to discover “regions” that interweave in complex manners, and, in the case of humans, that extend beyond the strict confines of the body into the socio-linguistic register” (1991, 102). It is certainly true that levels of value generation can be in conflict: how can it be that your body will fight for its life despite the deliberate attempt to end autopoiesis through an overdose of sleeping pills? Or, the other way around, how can the body attack itself in an autoimmune disease, to the dismay of the self that is able to express itself linguistically? Here, we disagree with Weber (2003), who seems to imply that value is always primordial and one-dimensional, that is, that everything that is of value to an organism can be ultimately derived from metabolic value (he calls it “existential value”). Such reductions may provide adequate description for

forms of life that do not involve high degrees of mediacy, but not if several levels of value generation coexist and come into conflict. For a smoker, the mechanisms of addiction may be explained with reference to metabolism, but it does not follow that smoking is in any way *about* survival in the way that breathing is.

How do different—sometimes competing, self-sustaining, and precarious—processes, spanning various levels of identity generation, often exceeding the boundaries of the autopoietic individual, relate to the cognizing subject? Could there be genuine values without autopoiesis? These are big questions that remain to be solved. But it seems clear that drawing a box labeled “value” is an unsatisfactory answer to these complex questions.

2.3.3 Modeling Values

In this section, we want to discuss how to model values following the enactive view. We see a large potential for advancing the enactive approach through the adequate use of synthetic models. However, it is very difficult to avoid remnants of Cartesian ways of thinking that are concealed in apparently innocent modeling assumptions. Partially rehearsing previous arguments by Rutkowska (1997), we want to uncover such “lurking homunculi” in “value-system architectures,” a class of architectures that feature a local mechanism to assign values.

The term “value system” is taken from the theory of neuronal group selection (TNGS), mainly proposed by Edelman and others (e.g., Edelman 1989). For instance, Sporns and Edelman define value systems as neural modules that are “already specified during embryogenesis as the result of evolutionary selection upon the phenotype” (Sporns and Edelman 1993, 968) and that internally generate reinforcement signals to direct future ontogenetic adaptation. A value system for reaching would become active if the hand comes close to the target. However, the point made here is not limited to neuronal group selection but instead extends to any model that features a strict functional and structural division between behavior-generating mechanisms and mechanisms of value-based adaptation, which we refer to as “value system architectures” in the following.

In order to point out the difficulties that result from such a separation of value judgment (built-in, figure 2.2a) from value execution (ongoing, figure 2.2b), we present two examples of our own research in simulated agent modeling. The deliberately simple first set of simulation experiments is described in more detail in Rohde 2010 and illustrates the difficulties of embedding functional modules into an otherwise dynamic and embodied

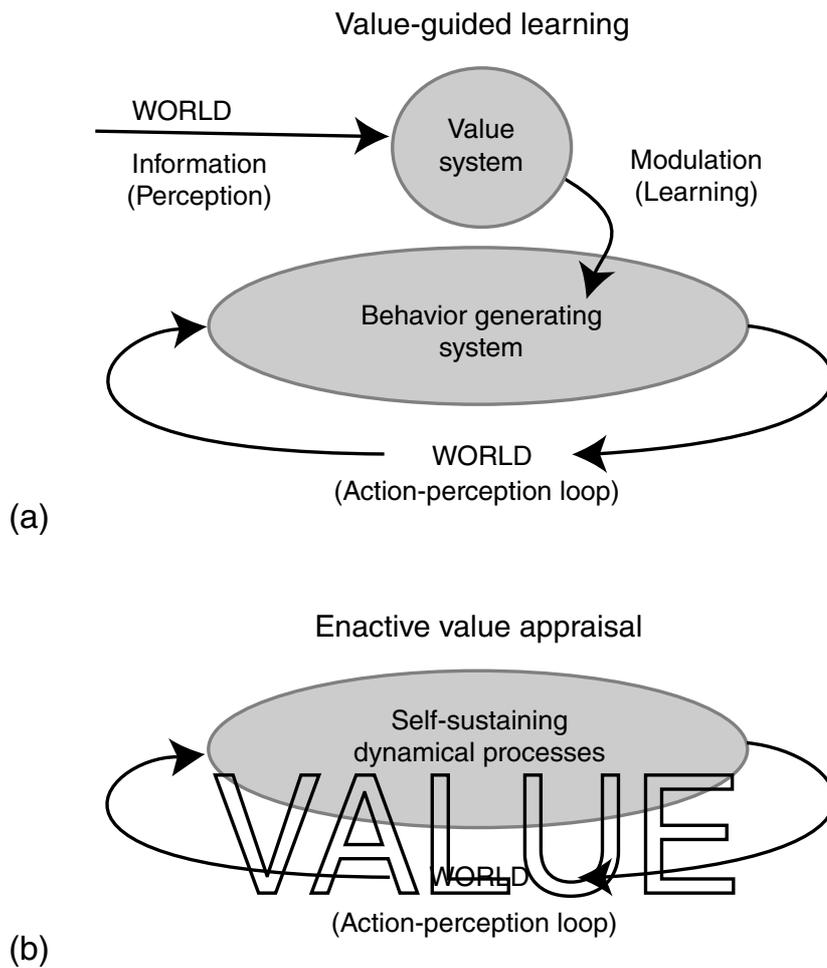


Figure 2.2

An illustration of the value systems (a) and the enactive approach (b) to conceptualizing values.

system. A mobile, two-wheeled agent is controlled by a neural network, which is generated automatically, using an evolutionary algorithm, such that the agent's behavior optimizes a performance measure. This "evolutionary robotics" technique mimics the principles of Darwinian natural selection in a simplified manner and is useful to the enactivist for several reasons. Because the performance criterion rates the *behavior* of an agent in a given environment, not its input/output mappings, this method provides a natural account of the situatedness, embodiment, and dynamics of behavior. Also, although the experimenter determines function by specifying the performance criterion, he or she underspecifies the mechanism that brings it about—this is shaped by automated search. Thereby, prior assumptions about the relation between function and mechanism are minimized, which can lead to behavior emerging from mechanism in ways that the

experimenter could not have foreseen, be it due to implicit prior assumptions, or due to cognitive limitations in dealing with complex dynamical systems (Harvey et al. 2005).

Value system architectures are inspired by findings on neural assemblies whose activity corresponds to salient events in the agent/environment interaction that are interpreted as internally generated reinforcement signals. In order to explore just how such a “value signal” could be generated, without caring yet about its function, an agent moving on a plane is evolved to perform light seeking behavior for a set of light sources presented sequentially and, at the same time, to generate a signal that corresponds to how well its approach to the light is being performed. Therefore, this value signal should go up only when the agent is progressing in its task.

The network controller evolved to control the two-wheeled simulated agent is extremely simple, but strikingly good at estimating how close the agent is to a light source, despite the poor sensory endowment (two light sensors generating only on/off signals) and the consequent massive ambiguity in the sensory space. The encircled group of three neurons is the part of the structure that generates the value signal (figure 2.3a). When investigating what this “value system” does, we find that it responds positively to activity on the left light sensor, but negatively to activity on the right light sensor, which, intuitively, does not make sense. The successful judgment can be understood only by taking the sensorimotor context into consideration, that is, the agent’s light seeking strategy (figure 2.3b). If the agent does not see the light, it turns to the right, until it senses the light with both sensors. It then approaches the light from the right, constantly bringing the light source in and out of range of the right sensor. In the end, the agent circles around the light source, perceiving the light with the left sensor only. Knowing this sensorimotor context, it is much easier to understand how the “value system” achieves a correct estimation of the distance to the light. The approach behavior starts only when the light is in range of the left light sensor, and this sensor remains activated from then on, which explains the positive response to left sensor activation. The right sensor, however, is activated only during the approach trajectory, and for increasingly short intervals, but not once the light source has been reached, and therefore is negatively correlated to progress in performance.

This simple example demonstrates an important theoretical possibility: a value signal that correlates to behavioral success, even if it is generated by a neural structure that is disconnected from the motor system, can rely

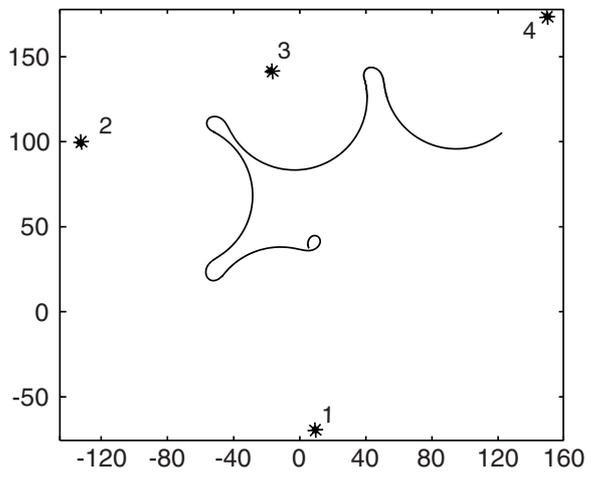
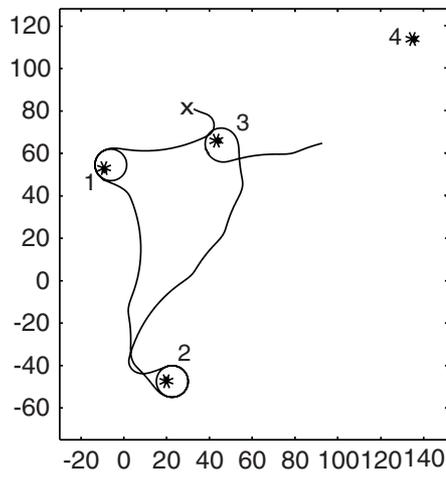
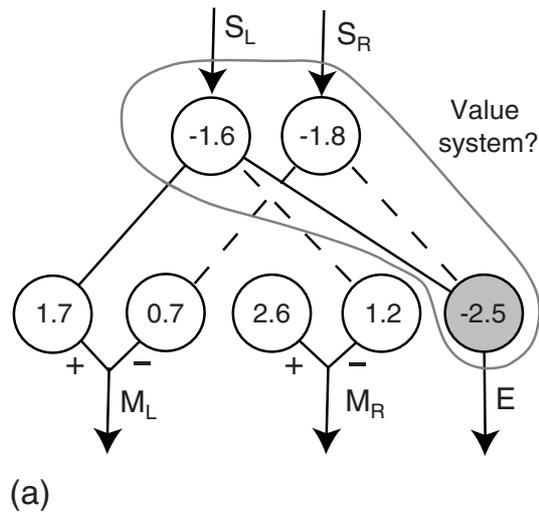


Figure 2.3

(a) The value-judging and light-seeking agent controller. (b) The successful light-seeking behavior. (c) The deterioration of light seeking through applications of the principles of neural Darwinism.

on an existing sensorimotor context. Why is this possibility important? Because it undermines the very idea of top-down behavioral adaptation on the basis of value system judgment: by identifying a correlation between activity in a separate cell assembly and behavioral success, we infer that this module is a value system (figure 2.2a) that informs the organism if a performed behavior is successful. But what if this module relies itself, in a circular fashion, on sensorimotor invariants in order to perform its judgment?

We explore this question in a second experiment. We allow the synaptic weights between sensors and motors (behavior generating subsystem) to change in order to maximize the output of the value system from the previous experiment. Such “neural Darwinism” is proposed as the source of adaptation in TNGS. We literally search for parameters of the behavior generating subsystem that make the agent behave so as to optimize the value signal. According to TNGS, this should result in an improved performance by the agent. In fact, figure 2.3c illustrates how, on the contrary, for this embodied value system, this type of parametrical modulation quickly results in a *deterioration* of performance. In a system that exploits sensorimotor couplings to generate a value signal, if these couplings are modified, their semantic contribution to the generation of meaningful judgment is gradually withdrawn, and we observe a semantic drift of the value signal: activity in the value system causes a change in behavior, which in turn causes a change of “meaning” of the activity of the value system, which causes a change in behavior, and so on. The system described previously, in isolation, rewards activity of the left sensor and punishes activity of the right. So if the semantic contribution of the sensorimotor couplings is gradually modified, the agent ends up avoiding the light source in a large circle, because this is the behavior that optimizes value system output, but not phototaxis.

This deterioration of performance is hardly surprising, given the structure of the value system and the way it works. But it demonstrates that value system architectures as outlined are not guaranteed to work without taking on board further premises. It has to be ensured that a value system estimates performance independently of the presence of reciprocal causal links, feedback loops, and semantic drift of local structures. If a value system is implemented in a rigid context, as it has been done in some robots with a limited behavioral domain (Verschure et al. 1995), the meaning of the signal can be preserved as independent of the modulation of behavior, such that the proposed circuits of adaptation do indeed work. However, in order to be convincing as a biological theory, it is necessary to specify how such a rigid wiring and “disembodiment” of value systems

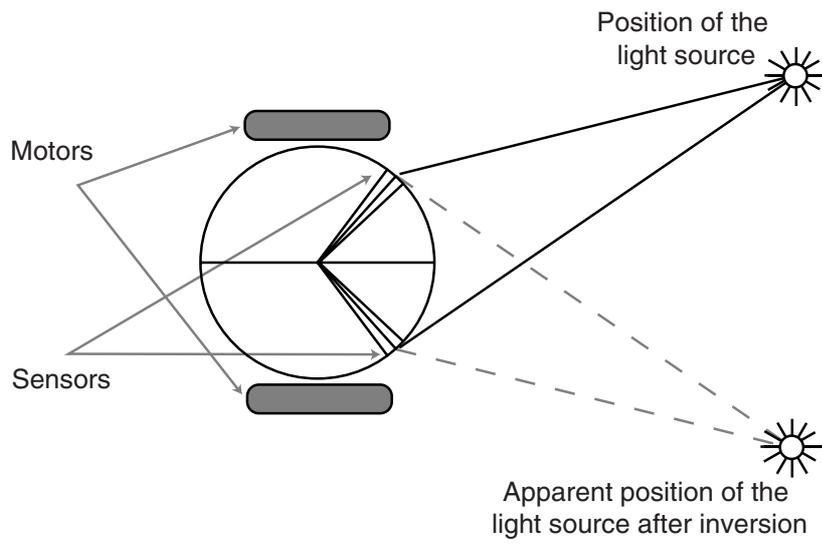
is realized in a living organism that is in constant material flux and embedded in multiple reciprocal loops with the environment. This is exactly the kind of problem that classical computationalist approaches have failed to answer satisfactorily. Indeed, we see value systems, because of their disembodied nature and top-down supervision of adaptation, as leftovers from a Cartesian mode of thinking. Such leftovers are not surprising; decades of exercising a computationalist methodology persist in the very language used to formulate questions.

An enactive approach, however, is based on the idea that values self-organize and emerge from a constantly varying material substrate. They are not reduced to local physical structures, such as a value system, and therefore there are no problems of explaining the semantic rigidity of material subunits.

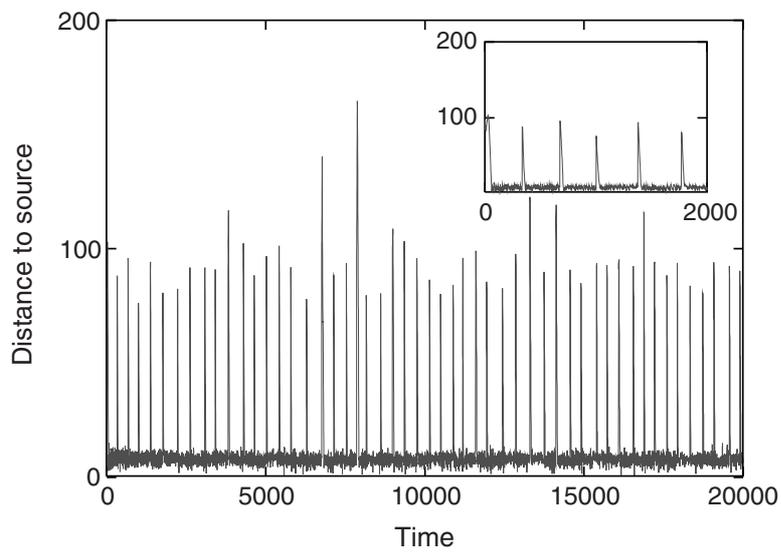
We now discuss an evolutionary robotics experiment that we conceive of as a first step toward a model of sense-making (Di Paolo 2000b). The task and agent are similar to the experiment described earlier, that is, seeking a sequence of different light sources (see figure 2.4a for a sketch of the agent). The controller consists of a network of homeostatic units, that is, neurons that regulate their connections to other neurons so that their own activity is maintained within a target range. This regulation is achieved by inducing local changes in the weighted connections, a design that is inspired by Ashby's homeostat (1960). These networks were set up to achieve both phototaxis and internal homeostasis by artificial evolution.

Every displacement of the light source (peaks in distance) is followed by a quick approaching behavior (figure 2.4b). The interesting fact about this agent is that it adapts against left/right swapping of its sensors, a situation that it has not been explicitly evolved to cope with (figure 2.4c): even though initially, the agent moves away from the light source—as we would expect if the visual field is inverted—over time it changes its behavior back to approaching the light; that is, the agent *reinterprets* its sensory channels according to the alterations of sensorimotor coupling it experiences, even though it had never been subjected to such alterations during evolution.

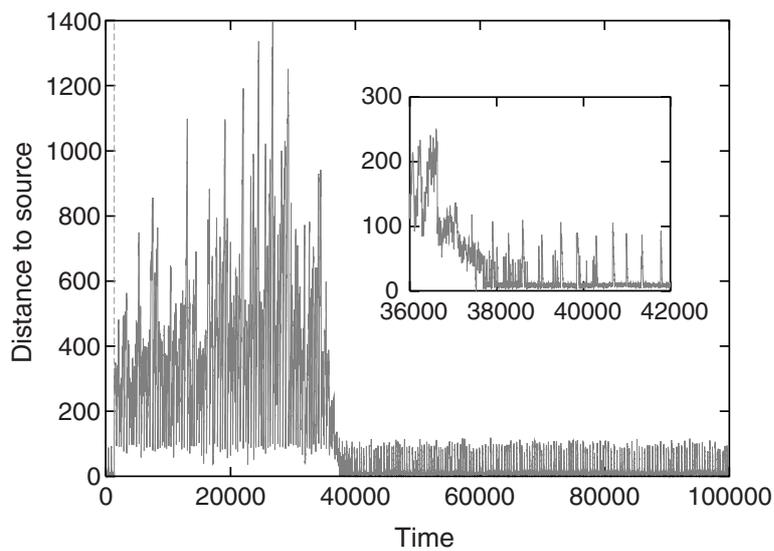
To what extent can these experiments be seen as more enactive than value system architectures? First, we ask: why does adaptation to visual inversion occur at all? Like in Bach-y-Rita's tactile vision system, there is no previous evolutionary history to explain how appropriate sense is made of a novel situation. Internal homeostasis acts as a dynamical organization trying to conserve itself, a minimal case of a self-sustained identity. The changes thus introduced can be said to conserve the *autonomy* of the neural process (see also Di Paolo and Iizuka 2008). This conservation has been linked through evolution to behavioral performance, that is, phototaxis.



(a)



(b)



(c)

Hence light is of positive value for this agent. When the body is disrupted, performance is disrupted as well, which can only be “interpreted” by the autonomous dynamics as a challenge to its conservation. The recovery of homeostasis results also in the reinterpretation of the sensorimotor coupling (and eventually in the regaining of phototaxis). However, the positive value of light demonstrated by the adaptive process cannot be reduced to the local plastic dynamics; it *emerges* through the ongoing internal and interactive dynamics of the agent in its environment. The meaning of light sensor activity and its functional role for phototaxis is dynamically constructed during the interaction. This minimal dynamic *sense-making* is very different from the a priori semantics of value systems we illustrated in our first model, which have to be protected from semantic drift. We find stability of neural dynamics, even if the system is not explicitly designed to serve as adaptation mechanism for a particular class of predicted problems, and this emergent meaningful adaptation can be explained through the study of mechanism and the parallel study of the behavior it brings about.

This example also demonstrates the usefulness of simulation modeling as a method for the enactive framework by showing how problems of functional reduction can be avoided, and even some degree of dynamical autonomy can be achieved that brings about adaptation through emergent value generation.

2.4 Enacting Social Meaning: Participatory Sense-Making

In this section, we explore what an enactive approach to social understanding would look like. Some authors have suggested ways of conceptualizing social understanding that touch upon some of the enactive ideas outlined in section 2.2. For Gallagher (2001), for instance, the basis of social understanding lies in the abilities of primary intersubjectivity (see also Trevarthen 1979). These include intentionality detection, the detection of eye direction, imitation, the perception of emotion, and meaning in postures

◀ Figure 2.4

Experiments in homeostatic adaptation using a two-wheeled light seeking agent (a). The agent’s distance to a long series of light sources is plotted as a function of time both for the case of normal (b) and inverted (c) visual fields. In (b), the agent approaches each new source of light that replaces the old one; in (c), immediately after sensors are inverted, the agent moves away from new light sources in its vicinity until adaptation ensues and light seeking behavior is recovered.

and movements. Thompson (2001) has suggested that we understand each other as part of an ongoing “self-other codetermination” that takes place when we are in interaction. But these approaches can be elaborated further. As with the case of values, the one important question that needs an answer before we can say that we grasp social understanding is: *where does meaning come from?* Current mainstream representational approaches do not give an answer to this question, nor do many embodied alternatives. The enactive approach, as we have seen, offers the concept of sense-making to address this issue. This section explains the kind of specific answer we can expect from an enactive approach to social cognition. One thing is clear: *interaction dynamics* as well as a specific notion of *autonomy* will play a crucial role in it.

2.4.1 Toward Enactive Social Understanding

Before laying out our proposal for an enactive approach to social understanding, let us examine the gaps in traditional takes on social cognition. The underlying assumption of central paradigms such as Theory of Mind theory (ToM) and simulation theory is that minds are enclosed and opaque, and hence others are puzzles for us to solve. The proposal of ToM is that we cognitively figure out others: we understand others by applying a capacity to draw logical inferences to sets of knowledge and perceptions. Simulation theory was proposed in reaction to what is thought of as the “cold reasoning” of ToM. We find out about what another is thinking or doing through an internal simulation of their behavior. Simulation comes in roughly two guises. There is Gordon’s radical simulationism, in which we act out the other’s stance (Gordon 1996), that is, we “become the other” for a short while in order to understand her. On the less radical version of simulation, we imagine ourselves in the shoes of the person we are trying to understand. The different versions of ToM and of simulation theory all presuppose a thorough disconnection between subjects and an internalist (thus hidden) view of intentions. In a social situation, we are confronted with an impenetrable other and so we find ourselves again and again thrown upon our own resources of reasoning and/or imagination.

Apart from suffering from internal contradictions, in this kind of approach the body plays no role of any significance. Issues of autonomy, emergence, and self-organization also remain untouched. As regards sense-making, meaning is supposedly derived from good old-fashioned information processing. Experience could be said to come into simulation approaches, but we would have to wax very lyrical about it—too much so—for the kind of experience implicated here to be anything like what it is understood to mean in an enactive approach.

Alternatives to both mind-reading and simulation approaches have been suggested. Gallagher, for instance, has criticized both because of their assumption that minds are private. Instead, he says, what we think, intend, desire, and so on is practiced, expressed and recognized in our body (Gallagher 2001, 2005). According to Gallagher, “in most intersubjective situations we have a direct understanding of another person’s intentions because their intentions are explicitly expressed in their embodied actions, and mirrored in our own capabilities for action” (2005, 224). Basically, we know others because of our own embodied experience—not so much because their bodies look like ours, but because we experience them as other persons through our own bodies. We are not confronted with an object to dismantle, but with someone that we already relate to at a very basic, bodily level. But proposals of embodied social cognition like these have a drawback: they often presuppose coupling between persons. Because of this, *how people interact* does not in itself become an explicit topic for investigation.²

But if we are to investigate social understanding along enactive lines, we need to pay special attention to the process of social interaction (De Jaegher 2006). Therefore, we suggest that, in order to understand social cognition, the embodied aspects investigated by several researchers need to be supplemented by an investigation of social interaction, in analogy with the interaction between agent and world described in section 2.2. In order to fully understand how meaning comes about in social understanding, we need not only to focus on the embodiment of interactors, but also on the interaction process that takes place between them.

There have been suggestions along these lines. Hobson (2002), for example, discusses “interpersonal engagement” or the intersubjective sharing of experiences that forms the fertile ground for the development of our capacity for thinking. There is also a large amount of research on dialog and interaction (see e.g., Sacks, Schegloff, and Jefferson 1974; Goffman 1983; Kendon 1990). This work has generated interesting results, but the research in these fields has often been geared toward notes on empirical findings, more than toward generating theoretical principles of communication or interaction. In order to start providing the latter, we need to look more concretely at the mechanism of social interaction as such.

2.4.2 Interaction and Coordination

An enactive approach to social understanding starts from the study of interaction and coordination. *Interaction* is here understood as the coupling between an agent and a specific aspect of its world: another agent.

Interaction is the mutual interdependence (or bidirectional, co-regulated coupling) of the behaviors of two social agents. Precisely which behaviors of the agents are implicated in this process will depend on the specific interaction and the situation in which it takes place (and on what its observer is interested in). What is of most interest right now, however, is what kinds of interdependence can exist.

In dynamical terms, systems can be *correlated*; that is, we may find similarities or coherences of behavior above and beyond what would be expected from what is known about their normal functioning. Of all the correlated behaviors, some are *accidentally correlated*, and some are *nonaccidentally correlated*. We are most interested here in the latter form of interdependence, which we call *coordination*. In social situations, coordination thus refers to the nonaccidental correlation of behaviors of two or more social agents. It is brought about by one or more common and/or connecting factors.

Imagine two people walking down the street. Suddenly, both of them turn their heads. Suppose we notice that their head-turning behavior has been prompted by someone screaming behind them. Their behaviors are thus *externally coordinated*, because there is a common external triggering factor. In the absence of such a factor, their behaviors might have been in a fortuitous correlation or the result of *precoordination*. When two people turn their heads at the same time because they are both—say, for some strange neurological reason—set to turn their heads every hour on the hour, the observed coherence is brought about by a preestablished coordination: their shared predisposition for hourly head turning. Again, there is a common factor: an internal “head-turning clock.” Common factors in precoordination can be of diverse origin, but often they involve a similar internal mechanism or shared histories.

Some precoordination is present in almost all social encounters, even if only by the existence of a common cultural background. But hardly any encounter, even with some precoordination, can unfold on the basis of it alone. Interactors need to coordinate their actions there and then. Such on-the-spot coordination is mostly achieved *interactionally*. We therefore call it *interactional coordination*, to refer to the fact that the interaction process itself plays a generative and facilitative role in the coordination. To illustrate, ways of greeting vary greatly between cultures, but even people with a common background may have to coordinate their hello’s and goodbye’s on their first rendezvous.

Moreover, coordination can also make interaction more likely to happen and continue. An example of this is making an appointment in order to

meet. Coordination thus can have an interactional function. This we call *functional coordination*. A beautiful example of this is the case of wolf circling (Moran, Fentress, and Golani 1981). Sometimes, as a wolf walks past another one that is seated, the sitting one gets up and starts to move in the opposite direction. However, rather than pass each other and walk away, they start to move in a circle together, head to tail. This behavior makes it possible for the wolves to size each other up, as it were, and to decide upon fighting or not, which can be said to be the function of this bodily coordination. Such coordination often serves an interactional function, namely that of facilitating or continuing the interaction, whatever it may lead to or change into.

Interactional coordination and functional coordination are not easy to separate; they are two sides of the same coin and describe the reciprocal influence between coordinated behavior and interaction as a process. As an extreme case of coordination through interaction we find the phenomenon of *one-sided coordination*. This happens when an individual coordinates *to* rather than *with* another. This distinction is further illustrated in the models described shortly.

In the following section, we discuss two examples of how to investigate these phenomena. We describe two evolutionary robotics models, one of which is based on an empirical study of “perceptual crossing.” Following this, we connect to the issue of meaning generation in social interaction via the introduction of the notions of *interaction rhythm* and *participatory sense-making*.

2.4.3 Modeling Embodied Coordination

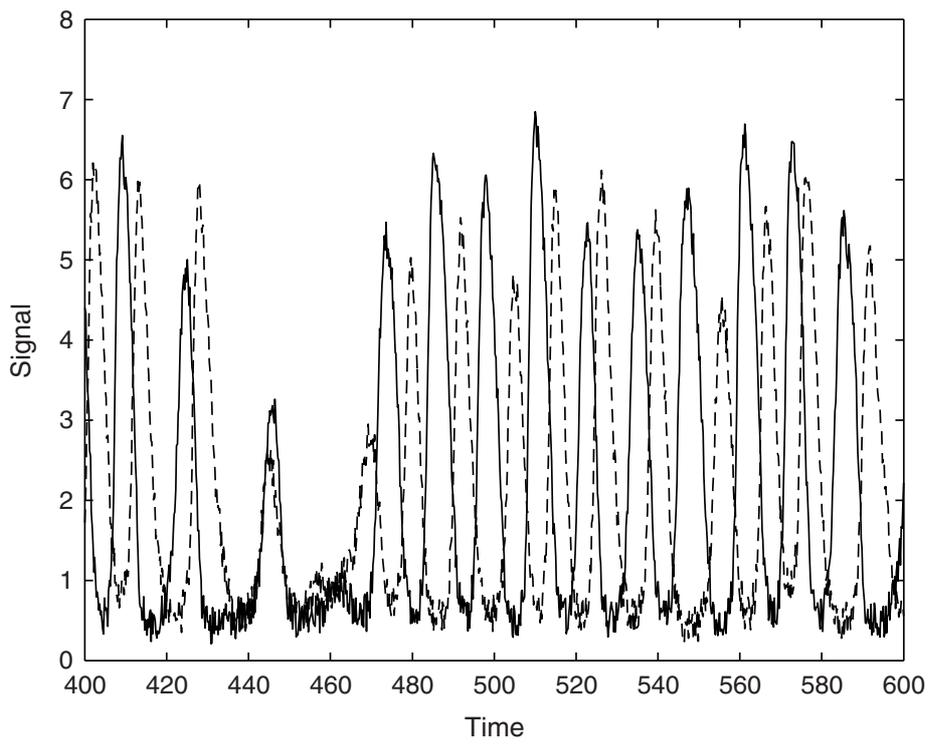
One approach to the question of how coordination between social interactors may be established is illustrated by some evolutionary robotics work on social interaction. More than half a century ago, W. Grey Walter explored simple forms of social robot coordination with his “tortoises” (Walter 1950). Such experiments demonstrated how a couple of very simple individual behaviors (such as wandering around and approaching a source of light) could result in complex, dance-like coordination when two such robots were put in mutual interaction. Recent studies using evolutionary methods also demonstrate this. For instance, in a simple simulation model, mobile agents must interact through an acoustic medium (Di Paolo 2000a). This work shows how different kinds of coordination are a direct result of the embodied interaction between agents over time.

The model again is deliberately simple. Two mobile agents are placed in an unbounded two-dimensional arena. Their bodies are circular and can

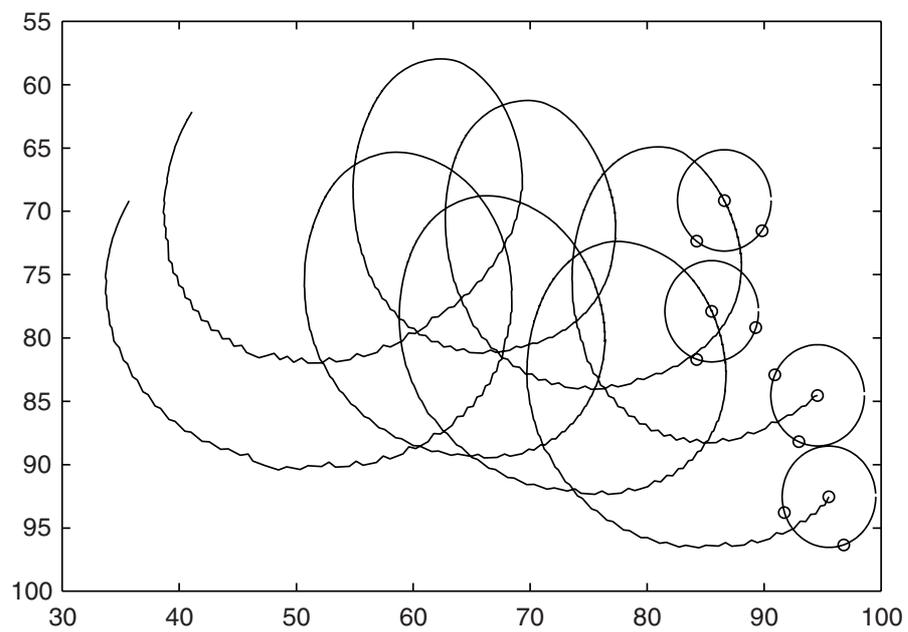
move by differential steering of two opposing wheels, which are controlled by a small continuous-time neural network. These agents are also provided with a loudspeaker that they use to regulate continuously the volume of the sound they emit. As sensors they have two microphones located symmetrically in their bodies, which are used to pick up any sound in the environment, including the sound they produce themselves. There is an inherent problem of distinguishing a signal produced by an external source and by the agent itself, because all sound signals are added up. A sound signal that travels through the body of an agent decays in intensity, so there is a significant difference to the sensor activity if the sound impinges directly on it or must go through the listener's body first; this self-shadowing property is indeed used by many mammals to detect sound source location.

With this setup, the task set for the agents is to locate and remain close to each other. There are no other restrictions to the agent's activity: they are allowed to evolve any kind of continuous sound signal or move in any way. The problem is nontrivial, because of the lack of other sensors and the single sound channel. Shouting at the top of their voices will not work, because the self-produced signal will overwhelm the sensors, but remaining quiet will not give any clue as to the agent's position that can be used to achieve the task. Consequently, sound must be used strategically. Because of their random initial positions, coordination between the agents must be achieved in order to facilitate a continuing interaction.

Successful agent pairs acquire a coordinated pattern of signaling in which individuals take turns in emitting sound so that each may hear the other. They solve the "self/nonself" distinction problem by making use of the self-shadowing property. If an agent constantly rotates, an external source of sound produces a regularly rhythmic pattern in the agent's sensors, while the sensing of its own signal is unaffected. A simple embodied strategy simplifies what would otherwise be a complex pattern recognition problem. This regular pattern affects their own sound production so that they also signal rhythmically, and finally through a process of mutual modulation the production of sound is coordinated in an anti-phase entrainment of signals. Further coordination is observed during interaction in proximity when patterns of regularly alternate movements are produced that resemble a dance (figure 2.5). Both the sound and movement coordination patterns are achieved through a process of coadaptation—tests on individual agents show that they are not capable of producing any of these behaviors in the presence of a noncontingent recording of a partner from



(a)



(b)

Figure 2.5

Sound patterns of agents in coordination: showing turn-taking activity (a), and trajectories of agents in coordination (b).

a previous interaction; that is, they are capable of interactional coordination but not of one-sided coordination (the same result has been found in models of detection of social contingencies; see Di Paolo, Rohde, and Iizuka 2008).

This and similar models demonstrate that the achievement of coordination *through* the interaction process is indeed something that can be expected to happen (as opposed to something that demands purpose-built mechanisms) in a broad range of dynamical systems in interaction. The agents in this example use their bodies and the time structure of their own movement to generate coordination. The generated patterns themselves help maintain the continuous coordination, and periods of breakdown followed by recovery of coordination are observed.

Experiments like these are sometimes disregarded because they seem so simple and “low-level” that it seems hard to see how they relate to human cognition. An alternative challenge for synthetic modeling is to try to account for empirical research conducted on human subjects that is driven by a similar aspiration for minimalism.

Auvray, Lenay, and Stewart (2009) have investigated the phenomenon of “perceptual crossing” in a similarly minimalistic manner. Blindfolded human participants interacting in a shared minimal virtual environment are asked to recognize the presence of each other. The only possibility to act is to move the cursor left and right along a virtual “tape” that wraps around using a computer mouse. Subjects sense the presence of an object or the other player only through a touch sensor whenever their own cursor “steps” on them. To make the task nontrivial, there is also a static object of the same size as the other subject on the tape (fixed lure), as well as a mobile object that shadows the motion of the other subject at a constant distance (attached lure). The problem is therefore not only distinguishing moving from nonmoving entities along the tape using the touch feedback, but distinguishing between two entities that move exactly the same, only one of which represents the “sensing” position of the other subject. Moment-to-moment sensory patterns therefore do not suffice to distinguish the three entities that may be encountered. Even so, recognition still results from the mutual search for each other.

Successful recognition relies on the global pattern of sensorimotor coordination between the participants, rather than on an individual’s capacity to express a confident judgment on whether a stimulus is caused by the partner. When subjects encounter a stimulus, they tend to oscillate around it, and these scanning movements only remain stable in the case that both players are in contact with each other. A subject could be fooled by the other player’s attached lure, but only to the point that the other player

remains largely on the spot (one-way coordination). This situation is unstable, as the other player will eventually move away to continue the search. Only when the two-way interaction condition is established does the situation remain globally stable. Hence the solution is truly interactional, because it is established by both partners searching for each other but does not rely on individuals performing the right kind of perceptual recognition between responsive and nonresponsive objects.

We have applied the technique of evolutionary robotics to gain further insight into this task (Di Paolo, Rohde, and Iizuka 2008). The virtual environment and task are the same and the agents are controlled by a neural network. The resulting global strategy is similar,³ but it raises an interesting further possibility regarding the role of interactional coordination. The empirical study shows that humans do not confuse a static lure with another subject. At first sight, it seems obvious that telling a mobile stimulus from a static one is the easiest task to solve in this experiment. Humans could, for instance, rapidly learn to discount changes to stimuli generated by their own movement using proprioception. The agents evolved in our model have another solution to the problem. If we take a closer look, we find a striking similarity between sensorimotor patterns for perceptual crossing involving the other player and for scanning a fixed lure (figure 2.6a, b). Encountering any stimulus makes the agent revert its direction of movement, which leads to another encounter followed by another inversion of velocity, and so forth. When we inspect the duration of the stimulus upon crossing a fixed object, we realize that it lasts longer than when crossing a moving partner. This is because the fixed object does not move itself. Therefore the *perceived size* differs for the two cases: longer in the case of a fixed object and shorter in the case of a moving object. The agent seems simply to rely on integrating sensory stimulation over time to make the distinction. This can be confirmed from the fact that the agent is quite easily tricked into making the wrong decision if the size of the fixed lure is varied.

What is interesting is that the smaller perceived size in the case of perceptual crossing depends on encounters remaining in an antiphase pattern (figure 2.6a). In other words, it depends on interactional coordination. Hence a systematic distinction in individually perceived size (between objects having the same objective size) is *co-constructed* during coordinated interaction, and in turn, individuals respond to the apparently smaller object by remaining in coordinated interaction. Looking more closely at the empirical data, Auvray, Lenay, and Stewart (2009) found that human participants may indeed use this strategy unconsciously to decide when to click. Here we see the importance of simple models as generators of ideas.

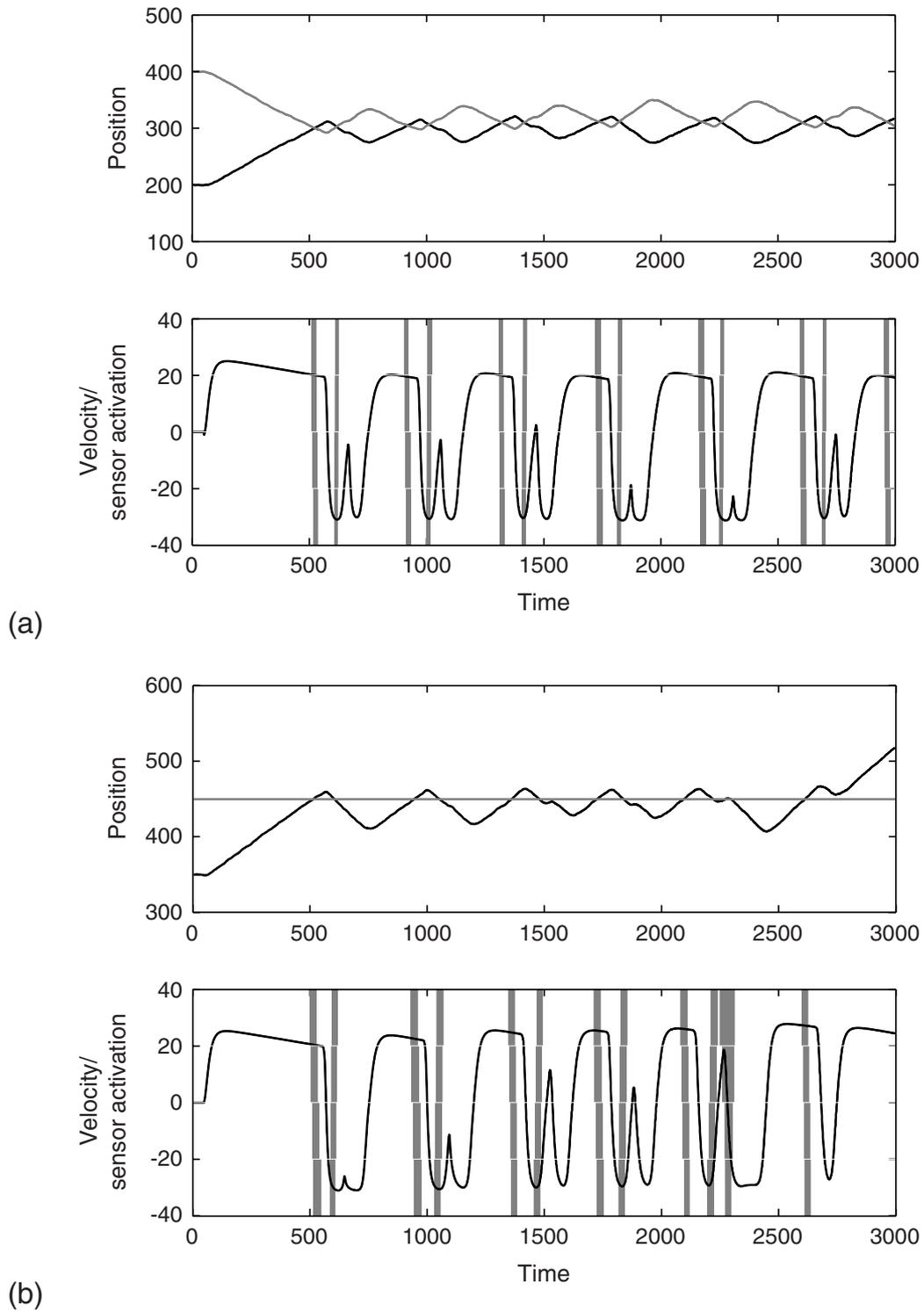


Figure 2.6

Perceptual crossing model. Top plots show the trajectories of agents over time; plots at the bottom show the motor commands (dark line) and sensor input (gray line). Stabilized social perceptual crossing (a); scanning of a fixed object (b).

These examples demonstrate the potential of an enactive modeling approach for the study of social interaction; instead of limiting the view to what happens inside one individual, the interaction process can be taken seriously, operationalized and studied in dynamical terms. Thereby these models have the possibility to capture the rich dynamics of reciprocity that are left outside of traditional individualistic approaches. The models demonstrate the importance of *timing* in interaction and suggest how it can affect sensorimotor processes at the individual level to the point that recognizing an interaction partner is possible thanks to the interplay and mutual modulation between the interaction and individual cognitive properties.

2.4.4 Social Perception as a Social Skill

How do we get from here to meaning generation in social encounters between humans? How do interactors understand each other? We believe that meaning generation and transformation can take place in the processes of interaction and coordination, as has also been suggested by the experiments discussed. Interactional coordination and functional coordination can be seen as the processes by which social encounters self-organize. In social situations in the human world, meaning is generated continuously in the interaction out of this self-organization, in combination with the histories, backgrounds, expectations, thoughts, and moods of the interactors. But how?

Our proposal is that enacting the social world relates to the precise timing of the functional and interactional coordination processes taking place in social situations. We call this timing *interaction rhythm*. Interaction rhythm refers to the diverse aspects of the temporality of the interaction—a necessary, though not sufficient, aspect of establishing, maintaining, and closing social interactions. Timing coordination in interaction is done at many different levels of movement, including utterances, posture maintenance, and so on. Rhythm as a term is preferred over the more general “temporality,” because it captures the *active* role that these elements play in the generation and organization of social interactions. As used here, the term “rhythm” does not refer to a continual strict temporal regularity or periodicity (the everyday meaning of the term), but rather to the *possible and actual temporal variability of timing in interaction*, including, at times and at certain levels of behavior, regular timing.

Interaction rhythm, moreover, refers to the self-organization in time of several elements and processes that *span* the individuals, that is, the organization of elements across and *between* individuals. This process can take

on a strong role and momentum of its own; it can itself become an autonomous phenomenon. In order to illustrate the potential autonomy of the social interaction, imagine the situation of two people walking in opposite directions in a narrow corridor, each on their way to their respective office. When they get close, they attempt to pass by each other. Each of them steps out of the way, but to the same side, so that they end up in front of each other and unable to walk on. And then they step to the other side, at the same time. And again. And again. They remain in front of each other for a brief time, meanwhile unable to continue on their way. Here, an unintended coordination takes place, and the interaction emerges and subsists—even if only briefly—outside of the intentions and goals of each of the two people.

If the interaction process is like this – sometimes autonomous and overriding or even working against individual intentions – it can alter and have an effect on the behaviors of the individuals involved. Another example is the perception of object size in the perceptual crossing model. In human interactions, the individuals involved are autonomous themselves, and this makes for the complexity of social interaction. If we are to understand meaning generation or sense-making in social interaction, we need to grasp what goes on in this interplay between the different states of the interaction process itself and those of the individuals engaged in it. We operationalize this interplay in our definition of social interaction: *Social interaction is the regulated coupling between at least two autonomous agents, where the regulation concerns aspects of the coupling itself and constitutes an emergent autonomous organization in the domain of relational dynamics, without destroying in the process the autonomy of the agents involved (though the latter's scope can be augmented or reduced)* (adapted from De Jaegher and Di Paolo 2007, 493).

How can we conceptualize social aptitude in this framework? We propose that social skill depends on a “rhythmic capacity.” This is not a capacity strictly of an individual, but one that comes about in interaction and is changed by both the interactional process and the individuals involved. We define this central capacity of social cognition as *the ability to coordinate through the interaction with another person*. Through such plastic coordination, the rhythm of an interaction can be adapted to varying circumstances, changes in goals, moods, and so on. This capacity is crucially dependent both on the individual interactors and on the process of engagement that ensues between them in every interaction.

In this view, social perception is not about finding hidden intentions in the other but is based on the mastery of self-other contingencies

(McGann and De Jaegher 2009). These contingencies tend to be mostly in the domain of values, intentions, and affects. They are negotiated during the encounter itself. They arise from interlocking bodily, interactive, and cultural processes, and unlike most cases of object perception they are protean in character. The social skills that must be mastered for social perception are the skills of coping interactively with regularities that can, and often will, change unexpectedly. According to McGann and De Jaegher, social skill is a mastery of negotiation.

We cannot say who is in charge of the process of the interaction. Therefore, here again, interactions need to be studied as wholes, plus their histories. Social meaning generation relies on the coordination of individual sense-making. It relies on *coordination as a process*, not an outcome. That is, precise mutual attunement of sense-makings is not necessarily the goal of interacting. Rather, it is the process of coordination between actions involved in sense-making that contributes to people understanding each other.

2.4.5 Participatory Sense-Making

To conclude, we propose the notion of *participatory sense-making* for social understanding in an enactive framework. Participatory sense-making is the extension of the enactive notion of sense-making into the realm of social cognition once we have taken into account the foregoing discussion about the importance of interaction (De Jaegher and Di Paolo 2007, 2008; De Jaegher 2009). In sense-making, active coupling with the world brings forth a realm of significance. In a social situation, the active coupling is with another social agent. Social agents can be engaged in individual sense-making, but when they start interacting, their sense-making is modified in accordance with the specific aspect of the world they are now interacting with—another social agent—according to the specifications laid out previously. Generation of social meaning relies on the process of coordination of individual sense-making, achieved in the interaction rhythm and by the rhythmic capacity.

Not only this, but participatory sense-making also opens up domains of sense-making that are not open to the individual on his own. Participatory sense-making constitutes a continuum from less participatory to highly participatory sense-making. At the former end of the spectrum, we find for instance *orientation*, in which individual A orients B to aspects of B's cognitive domain. This is not very participatory, because there is not much mutuality to the sense-making. As we move away from this end of the range, the sense-making activities of the individuals involved are

increasingly mutually changed by their coordinated sense-making, and also change it. At the most participatory end of the spectrum, individuals truly intertwine their sense-making activities, with consequences for each in the process, in the form of the interactional generation of new meanings and the transformation of existing meanings. Academic collaborations are a good example of this. Sometimes, when the partnership is especially fruitful, a completely new vantage point on a problem arises, or a fresh interpretation of a result occurs. Sometimes it is impossible to attribute this development to one of the participants only.

It can seem as if the account we propose here is only applicable to live interactions. But what about observational social understanding—for instance, when watching a film? Or even understanding someone in a letter exchange or in an email conversation? How does social understanding happen when there is no live interaction and coordination? The point of our proposal is not that social understanding only happens in situations where the participants are physically present to each other. It is rather that social understanding has its developmental and logical origins in social interaction. Without experience of interacting in development, we would not reach the sophisticated forms of intersubjectivity we have. And without studying interactional coordination, scientists cannot get to the bottom of how it is that we understand others.

In conclusion, using the notions of sense-making, interaction rhythm, the rhythmic capacity, and our definition of social interaction enables us to conceptualize social understanding as something that is enacted—co-constructed—in the interaction. We do not need to posit a specialized module in the brain, but propose to explain social understanding as always based on and supported by the dynamics of interaction between the cognitive agent and the environment. Because an enactive approach places great importance on the autonomy of the individuals involved, this approach to social cognition, while focusing on the interaction process, paradoxically also gives social agents an autonomy and role that has not been thematized before: that of participation in contrast to mere observation.

2.5 Play: Enactive Re-creation

We come back to some of the problems raised in the introduction. This section will draw on what has been learned so far about the horizons of enactivism to approach the general question of human cognition (the umbrella term under which cognitive scientists gather conceptual thinking, planning, language, social competences, and so on).

We have already mentioned that the impact of the enactive approach in cognitive science, and that of embodied and dynamical views in general, has been acknowledged by many sectors, but not yet as a proper replacement for representationalism in what concerns higher level cognition. Some arguments have been advanced regarding the very possibility of a nonrepresentational framework for this task. Clark and Toribio (1994) question how the very situatedness of action-oriented and richly dynamical couplings between agent and environment is not at the same time responsible for “tying down” cognition to the present situation. Internal representations, the argument concludes, will have to reenter the picture to account for activities that seem decoupled from the current situation, such as picturing the house of your childhood.

The argument is right in that, from an enactive perspective, such high-level skills are indeed still unexplained. But the argument simply assumes that they are also unexplainable in enactive terms. Importantly, the argument relies on a misunderstanding of the idea of situatedness. To say that we are present in a situation with our bodies does not mean that the situation boils down to the physical couplings that we encounter, that is, that we are shackled to our present physical circumstances. This is why the concept of sense-making is so interesting. It is all too easy to interpret this idea in a one-sided manner—events in the world are given meaning by the agent—and ignore the crucial possibility that the cognitive agent may also be an *active creator of meaning* and that such creation can be subject to change and eventual control by emergent levels of cognitive identity.

Could this point be a way of making progress in an enactive account of human cognition? Let us try to formulate the essence of the problem first. What is essential to human cognition as opposed to other forms of animal cognition? Margaret Donaldson (1992) formulates the issue in a very useful way. She puzzles about the amazing human capability of constantly inventing new goals so that we invest them with value and submit passionately to them (sports, hobbies, record breaking). An explanation of human intelligence should perhaps not concentrate so much on issues such as, say, how we manage to do math. It should bring to the center the question of *why* we do those things at all—when did they become valuable for us?

Donaldson describes different ways to be a human mind. As a developmental psychologist, she concentrates on how transitions between these different modes occur throughout a lifetime. The question parallels how Jonas and others have treated the history of life and mind as transitions in scales of mediacy. Donaldson distinguishes four modes in which we

function as minds depending on the focus of concern. This is amenable to the whole of our previous discussion. To have different foci of concern is no more or less than to have different modes of value generation. The *point mode* deals with here-and-now coping (most animal activity, skillful practices in humans). The *line mode* expands the focus of concern to the immediate past and the possible future as well as to other spatial localities (understanding of immediate causes and consequences of events). The *construct mode* produces a decentering of cognitive activity; concern focuses on events that have happened or may happen at some point in time or somewhere, and not necessarily involving the cognitive agent (induction, generalization). Finally, the *transcendent mode* has no locus; it deals with nowhere, no-time (abstract thought, metaphysics).

These modes are manifested to different degrees in different circumstances and with respect to different mental “components” such as perception, action, emotion, and thought. The modes are transversed developmentally, building upon previous stages. The generation of different kinds of intention and the manipulation of our own consciousness are the central factors in this development. This backdrop can help us describe our problem as that of formulating an enactive account of how to move beyond the point mode and into the line and construct modes.

This transition indicates the development of a capacity to “unstick” meanings from a given situation and “stick” novel ones onto it (to put it graphically), or, generally, the capacity to influence meaning generation. This has confusingly been described as offline intelligence (Clark 1997; Wheeler 2005), whereas “decentering” or “meaning manipulation” may be better labels. Such a capability is indeed a challenge for dynamical accounts of cognition that emphasize coupling with the environment. It would seem that cognitive activity is “glued” to the here and now in such accounts, that is, always in the point mode. By contrast, cognitivism sees no challenge in this. Manipulation of representations to deal with the here and now is not fundamentally different from manipulation of representations to deal with the there and then, or with nowhere, no-time. This is hardly surprising. Cognitivism starts at the high end of the spectrum. It is based on nontemporal, nonspatial, unsituated mechanisms (and consequently, its own challenge is how to move in the opposite direction, toward ongoing coping).

If we look historically or developmentally for an activity that could play a part in this transition we must conclude that (1) it should be an embodied activity, accountable for by means of the many skills that we can already explain in enactive terms, and (2) it should allow for ambiguity of

meaning as well as the generation of novel kinds of value. The worst possible candidates are concrete goal-directed activities in which meaning is well defined by situational constraints. The best candidates are those goal-generating activities where meaning is fluid. Jonas points to image-making, which is indeed an excellent example. But it is already too sophisticated and immediately invites representational thinking. More parsimonious possibilities include dance, music, ritual, and play. Here we briefly explore the last of these possibilities.

Can we sketch an enactive account of play? There is a significant literature on play in animals as well as different forms of play in human children and how play relates to socialization, self-regulation, attachment, use of language, and the development of cognitive capabilities.⁴ The interesting fact for our present discussion is that elements of the meaning manipulation that this activity can afford are already present in all forms of play. We have already mentioned the possibility of sense-making leading to increasingly removed manipulation of meaning. Might not the presumed bacterium swimming up a saccharine (not sugar) gradient and the young baboon accepting to be chased around by the smaller playful infant share something in common? Are not both deceived to different degrees in their sense-making activities, the one unknowingly, the other willingly?

The first thing to note about play is that it is hard to define and easy to recognize. Miller (1973) lists some properties of play, such as the repetition of motor patterns, lack of economy, exaggeration, lack of a direct practical end, production of novel sequences of behavior, combinatorial flexibility, egalitarianism, and others. Play occurs only in the absence of more urgent motivations related to survival; hence it is the privilege of species where individuals have enough spare energy, time, and protection. Not all animals do it, and in those species that play, mainly infants and juveniles do it—exceptions are humans and species that are given safety through their adulthood such as cats, dogs, and domesticated monkeys and apes. Evolutionary explanations of play abound. They typically refer to beneficial by-products such as the training of motor skills. The merits of such explanations must be assessed in each individual case, but in general terms understanding even quite “unsophisticated” bodily play (rough and tumble, simulated pursuit-evasion, etc.) cannot be fully achieved without an experiential approach. Much is missed if we cannot understand why animals are *interested* in play. Maxine Sheets-Johnstone (2003) answers this question by indicating the dimension of kinesthetic feeling that animals explore in play: the dimension of corporeal powers, the *I-can* and *I-cannot*.

The experiential dimension of value explored in this way is opened up by the element of social interaction and the forms of participatory sense-making that it affords. It is here that kinaesthetic pleasure turns into make-believe. Running may be fun, running from or after someone even more. The excitement of aggressive or sexual encounters can be safely explored if distinguished from real ones by appropriate signals and conventions. It is this novel way of socially exploring the meaning of fake situations using real and concrete interactions that is taken to its pinnacle by humans in the form of pretend play. Here we are already at the other side of the transition, because if the arrangement of wooden cubes can be a house and the pen a spaceship, the root capacity of meaning creation and manipulation is already going strong.

Cognitivist accounts of pretence in play, such as Leslie (1987), go very much in line with similar accounts of social understanding already examined, and their criticisms, such as Hobson (1990), complete the parallel. Piaget's views on pretend play are closer to the enactive approach (1951). For him, the beginnings of play are rooted in the assimilative function, whereby new situations are coped with using existing sensorimotor schemas. A fifteen-month-old infant deals with a pillow using certain actions (touching, laying his or her head on it, going to sleep). As soon as another object (a blanket) is assimilated into the same structures, it becomes a make-believe pillow. The infant finds pleasure in the assimilative function and smiles. Donaldson (1992) criticizes this view (see also Sutton-Smith 1966). If only assimilation were taking place, the blanket and the pillow would be indistinguishable. There would be no reason to smile unless there was a simultaneous awareness of the difference between the two cases and the sense of "getting away with something." Make-believe relies crucially on the *combined similarity and difference* between two situations: one concrete, tied to physical events, and the other in terms of manipulated meaning (the tension of this combination reappears in other creative activities such as making images).

The view of play as predominant assimilation misses out on the active element of construction of new environmentally and bodily mediated meaning. Play breaks from the constraints of self-equilibrating cognition. It does not have the structure of a cognitive confrontation with an environment that places demands on the agent. Play is precisely *not* a problem requiring a solution. In fact, play is the breaking of this pattern, or rather its redeployment into an active construction of meaningful action where no such sense-making is directly demanded from the environment or from definite internal needs. The urge to play (at least during the creative phases

of play) is indeed present but remains undefined until the activity of play itself helps the child make this urge clearer.

How is this possible within an accommodation/assimilation/equilibration dynamic? It seems impossible if we resist acknowledging the active participation of the child in transforming his or her world. Vygotsky (1966) gives us a glimpse of how such manipulation of sense-making could happen. In play, the child begins to detach meaning from a situation and to regulate such meanings first with respect to objects and later to his or her own actions. This is motivated by the inability to satisfy immediate needs. Play becomes a way of substitution for real satisfaction and a way of dealing with an insurmountable mediacy. Soon the value-generating properties of play become evident and the activity is done for its own sake. "Detachment" is an embodied activity. It begins by relying on concrete similarities—a doll resembles a person—but soon these similarities are mostly given by the child's own use of gestural schemas and not the objects themselves (Watson and Jackowitz 1984). If something is treated as a horse, if it is made to move and sound like a horse, then the child accepts it as a horse (without forgetting that it is not one). This is the ambiguity that, according to Donaldson (1992), can produce laughter and exhilaration, the bringing into presence of what is not there, a cheating of "reality."

Once objects in the environment are imbued with meaning by actions that in turn demand from the child an (adaptive) interpretation, these objects become toys—would-be cars, houses, and creatures. The child is now acting at the pinnacle of his or her enactive capabilities, because he or she is bringing forth an alienated meaning through gestural schemas and then—and here lies the equally radical trick—submitting to the reality thus created through adaptive equilibration (the absence of which would make play unchallenging and "unreal").

The combination of a concrete embodied situation with alienated virtual meaning is the freedom-engendering paradox of play. But it would not be a paradox if all there was to pretence was the manipulation of internal representations. This would result in no sense of ambiguity. Cognitivism cannot explain fun. When the child becomes the regulator of play, the activity takes off as a proper form of life. The child explores the new freedom by following pleasurable activities, but at the same time, learns to generate new rules—new constraints that structure and reevaluate reality and that must be followed strictly (otherwise play becomes random and boring). The child is unhappy if he or she cannot bounce the ball more than the nine times managed so far. The norm is arbitrary, invented by the child, but in allowing his or her body to submit to it, it becomes

as serious as other social or biological norms. The player is the lawgiver and the rule-follower, the question maker and the responder. Play is thereby autonomous in the strict sense advocated by enactivism because it engenders a self-sustaining network of activities.

Pretending is only possible if a novel way of generating norms and values arises in conjunction with exploratory play. The best players are those that create new rules in a contextual manner so that they can continue to play and fun does not run down by exhausting the possibilities of the game. Rules are made up in play; they are solidified versions of norms. Fun is the exploration of the limits thus imposed on bodily activity and social interaction. But when the possibilities are extinguished, the game becomes boring. Fun is also the change and revision of norms that reopen play. Over time, play is a self-structuring process governed by the dialectics of expansion and exhaustion of possibilities. Its freedom lies in the capability that players acquire of creating new meaningful (not arbitrary) constraints. The playful body is a new form of autonomous being—a novel mode of the cognitive self. It can now steer its sense-making activity and set new laws for itself and others to follow. This might help to answer the question we raised at the end of our discussion about embodiment in section 2.2.4.

We find that play is an area particularly rich for the exploration of enactive themes from emergence of identities and levels of social coordination, to manipulation of sense-making through experientially guided bodily action. Perhaps no other framework is better placed to explain play and its paradoxes, and this may be why there is such a paucity of references to play in cognitive science. When a child skillfully supplements the perceptual lack of similarity between a spoon and a car by making the spoon move and sound like a car, he or she has grasped in an embodied manner the extent to which perception can be action-mediated. With his or her body, the child can now alter sense-making activity, both on external objects, as well as his or her own actions and those of others. The child has become a practitioner of enactive *re-creation*.

2.6 Conclusion

A proper extension of the enactive approach into a solid and mainstream framework for understanding cognition in all its manifestations will be a job of many and lasting for many years. This chapter has attempted only to point to specific directions and show that enactivism can be made into a coherent set of ideas, distinguishable from other alternatives, and that it

can provide the language to formulate problems and the tools to advance on issues that are sometimes out of the focal range of traditional perspectives (if not in their blind spot). The strength of any scientific proposal will eventually be in how it advances our understanding, whether in the form of predictability and control, or in the form of synthetic constructions, models, and technologies for coping and interacting with complex systems such as education policies, methods for diagnosis, novel therapies, and so on. For this, it is crucial for ideas to be intelligible and promising.

In this respect, we would like to draw attention to the valuable role played by minimal models and experiments. Their function goes beyond the study of a given phenomenon. Minimal modeling provides crucial conceptual training that would be hard to obtain otherwise (Beer 2003; Harvey et al. 2005; Rohde 2010). Analytical thinking is at home with linear causality, well-defined and unchanging systems, and reduction. The alternatives of emergent, many-layered, causally spread, nonlinear systems in constant constitutive and interactive flux are very hard to manage conceptually. This is an important focus of resistance to many enactive ideas. It is here that synthetic modeling techniques may have their major impact: in producing novel ways of thinking and generating proofs of concept to show that some proposals may not be as coherent as they sound (as in our critical study on value system architectures) or to demonstrate that apparently hazy concepts find clear instantiations even in simple systems (as in the case of emergent coordination through social interaction processes). Methodological minimalism is, therefore, a key element contributing to the acceptability of enactive ideas.

Models that attempt to illuminate the enactive framework must take into account the core ideas of enactivism. A serious take on embodiment will depend on the extent to which a system's behavior relies nontrivially on its body and its sensorimotor coupling with the environment as opposed to input/output information processing. Emergent properties and functionality will contrast with misplaced localization in subagential modules. Autonomy, to the extent that it can be captured in simulation or robotic models, will depend on how the model instantiates the dynamics of self-constituted precarious processes that generate an identity and how such processes create a normativity at the interactive level that leads to sense-making. Enactive modeling must also relate to experience. As a scientific tool, it belongs to the realm of third person methods, so the relation will have to find its place in the process of mutual constraining that has been proposed for the empirical sciences and first person methods already mentioned.

Alongside the explorations presented in this work and the horizons of questions, methods, and explanations that they open, there will be many other areas where enactive views could make a contribution. We repeat that we have not aspired to be exhaustive in neither breadth nor depth. But we do think that we have moved in the direction in which enactivism could grow the strongest: toward higher forms of cognition. Some of the ideas we have explored raise more questions than definitive answers. And this is as it should be in the current context. Focusing on the core concepts of enactivism has been a way of changing perspectives on well-known problems. This will inevitably lead to novel questions, which we have raised throughout the chapter. How do different modes of value-generation coexist in a human subject? How does sense-making get socially coordinated through different kinds of participation? How is the creation of novel meaning achieved in transitional activities such as play? Each of these areas indicates a direction in which much further work is needed and that might possibly lead to newer horizons.

Notes

1. Emergence in this view is close to the notion proposed in Thompson and Varela (2001) and Thompson (2007) with the exception that our second requirement is there presented only as a possibility. We favor a stronger definition, because we want to emphasize the role of mutual causation in order to introduce a sharper contrast between enactivism and reductionism.
2. This is a problem shared by other sensorimotor theories of social cognition such as those built upon the role of “mirror neurons” (Gallese 2001); additionally, such neural correlations themselves should be treated as suspect of the meaning reduction criticized in section 2.3.
3. Interestingly, the agent’s behavior resembles the human behavior only if we include a delay between an agent’s encounter of an object and input to the neuro-controller. If such a delay is not present, the agent’s position eventually converges to a fixed point and stands still. This result raises an interesting question: why do participants keep oscillating around each other, rather than to just “stand on top of each other” after recognition? Our model predicts that sensory delays play a role in this phenomenon and that the amplitude of the scanning oscillations around a target is positively correlated with the amount of delay.
4. Although there is a paucity of research on play strictly from within cognitive science, important relevant works on the subject can be found in the fields of cultural anthropology (Schwartzman 1978), developmental psychology (Sutton-Smith 1997), phenomenology (Fink 1968), animal behavior (Fagen 1981), psychoanalysis (Winnicott 1971), and social science (Goffman 1961; Huizinga 1949).

References

- Ashby, W. R. (1960). *Design for a brain: The origin of adaptive behaviour*. 2nd edition. London: Chapman and Hall.
- Auvray, M., Lenay, C., and Stewart, J. (2009). Perceptual interactions in a minimalist virtual environment. *New Ideas in Psychology* 27 (1): 32–47.
- Bach-y-Rita, P., Collins, C. C., Saunders, F., White, B., and Scadden, L. (1969). Vision substitution by tactile image projection. *Nature* 221:963–964.
- Barandiaran, X., Di Paolo, E., and Rohde, M. (2009). Defining agency individuality, normativity, asymmetry and spatio-temporality in action. *Adaptive Behavior* 17 (5): 367–386.
- Beer, R. (2003). The dynamics of active categorical perception in an evolved model agent. *Adaptive Behavior* 11:209–243.
- Beer, R. (2000). Dynamical approaches to cognitive science. *Trends in Cognitive Sciences* 4:91–99.
- Brooks, R. A. (1991). Intelligence without representation. *Artificial Intelligence* 47:139–159.
- Bruner, J. (1966). *Toward a theory of instruction*. Cambridge, MA: Harvard University Press.
- Chiel, H. J., and Beer, R. (1997). The brain has a body: Adaptive behavior emerges from interactions of nervous system, body and environment. *Trends in Neurosciences* 20:553–557.
- Chrisley, R. (2003). Embodied artificial intelligence. *Artificial Intelligence* 149:131–150.
- Clark, A., and Toribio, J. (1994). Doing without representing? *Synthese* 101:401–431.
- Clark, A. (1997). *Being there: Putting brain, body, and world together again*. Cambridge, MA: MIT Press.
- Clark, A., and Grush, R. (1999). Towards a cognitive robotics. *Adaptive Behavior* 7:5–16.
- Damasio, A. (1994). *Descartes' error: Emotion, reason, and the human brain*. New York: Putnam.
- De Jaegher, H. (2006). Social interaction rhythm and participatory sense-making. An embodied, interactional approach to social understanding, with implications for autism. Unpublished DPhil thesis, University of Sussex.
- De Jaegher, H. (2009). Social understanding through direct perception? Yes, by interacting. *Consciousness and Cognition* 18 (2): 535–542.

De Jaegher, H., and Di Paolo, E. (2007). Participatory sense-making: An enactive approach to social cognition. *Phenomenology and the Cognitive Sciences* 6 (4): 485–507.

De Jaegher, H., and Di Paolo, E. (2008). Making sense in participation. An enactive approach to social cognition. In *Enacting intersubjectivity: A cognitive and social perspective to the study of interactions*, ed. F. Morganti, A. Carassa, and G. Riva, 33–47. Amsterdam: IOS Press.

Dennett, D. C. (1993). Review of F. Varela, E. Thompson, and E. Rosch, *The Embodied Mind*. *American Journal of Psychology* 106:121–126.

Dewey, J. [1929] (1958). *Experience and nature*. 2nd edition. New York: Dover.

Di Paolo, E. A. (2000a). Behavioral coordination, structural congruence and entrainment in acoustically coupled agents. *Adaptive Behavior* 8:27–47.

Di Paolo, E. A. (2000b). Homeostatic adaptation to inversion of the visual field and other sensorimotor disruptions. In *From Animals to Animats 6: Proceedings of the Sixth International Conference on the Simulation of Adaptive Behavior*, ed. J.-A. Meyer, A. Berthoz, D. Floreano, H. Roitblat, and S. Wilson, 440–449. Cambridge, MA: MIT Press.

Di Paolo, E. A. (2005). Autopoiesis, adaptivity, teleology, agency. *Phenomenology and the Cognitive Sciences* 4:429–452.

Di Paolo, E. A. (2009). Extended life. *Topoi* 28:9–21.

Di Paolo, E. A., and Iizuka, H. (2008). How (not) to model autonomous behaviour. *BioSystems* 91:409–423.

Di Paolo, E., Rohde, M., and Iizuka, H. (2008). Sensitivity to social contingency or stability of interaction? Modelling the dynamics of perceptual crossing. *New Ideas in Psychology* 26 (2): 278–294.

Donaldson, M. (1992). *Human minds: An exploration*. London: Penguin Books.

Dreyfus, H. L. (2002). Intelligence without representation—Merleau-Ponty's critique of mental representation. The relevance of phenomenology to scientific explanation. *Phenomenology and the Cognitive Sciences* 1:67–383.

Edelman, G. M. (1989). *The remembered present: A biological theory of consciousness*. Oxford: Oxford University Press.

Engel, A. K., Fries, P., and Singer, W. (2001). Dynamic predictions: Oscillations and synchrony in top-down processing. *Nature Reviews Neuroscience* 2:704–716.

Fagen, R. (1981). *Animal play behavior*. Oxford: Oxford University Press.

Fink, E. (1968). The oasis of happiness: Toward an ontology of play. *Yale French Studies* 41:19–30.

- Gallagher, S. (1997). Mutual enlightenment: Recent phenomenology in cognitive science. *Journal of Consciousness Studies* 4:195–215.
- Gallagher, S. (2001). The practice of mind: Theory, simulation or interaction? In *Between ourselves: Second-person issues in the study of consciousness*, ed. E. Thompson, 83–107. Exeter, UK: Imprint Academic.
- Gallagher, S. (2005). *How the body shapes the mind*. New York: Basic Books.
- Gallese, V. (2001). The “Shared Manifold” hypothesis: From mirror neurons to empathy. In *Between ourselves: Second-person issues in the study of consciousness*, ed. E. Thompson, 33–50. Exeter, UK: Imprint Academic.
- Goffman, E. (1961). *Encounters*. Indianapolis: Bobbs-Merrill.
- Goffman, E. (1983). The interaction order. *American Sociological Review* 48:1–17.
- Goldstein, K. [1934] (1995). *The organism*. New York: Zone Books.
- Gordon, R. M. (1996). “Radical” simulationism. In *Theories of theories of mind*, ed. P. Carruthers and P. K. Smith, 11–21. Cambridge: Cambridge University Press.
- Grush, R. (2004). The emulation theory of representation: Motor control, imagery and perception. *Behavioral and Brain Sciences* 27 (3): 377–396.
- Harvey, I., Di Paolo, E. A., Wood, R., Quinn, M., and Tuci, E. (2005). Evolutionary robotics: A new scientific tool for studying cognition. *Artificial Life* 11:79–98.
- Harvey, I., Husbands, P., Cliff, D., Thompson, A., and Jakobi, N. (1997). Evolutionary robotics: The Sussex approach. *Robotics and Autonomous Systems* 20:207–224.
- Heidegger, M. (1962). *Being and time*. Trans. J. Macquarrie and E. Robinson. Oxford: Blackwell.
- Hobson, P. (1990). On acquiring knowledge about people and the capacity to pretend: Response to Leslie (1987). *Psychological Review* 97:114–121.
- Hobson, P. (2002). *The cradle of thought*. London: Macmillan.
- Huizinga, J. (1949). *Homo Ludens*. London: Routledge.
- Husserl, E. (1973). *The phenomenology of internal time-consciousness*. Ed. M. Heidegger, trans. James S. Churchill. Bloomington: Indiana University Press.
- Hutchins, E. (1995a). *Cognition in the wild*. Cambridge, MA: MIT Press.
- Jonas, H. (1966). *The phenomenon of life: Towards a philosophical biology*. Evanston, IL: Northwestern University Press.
- Juarrero, A. (1999). *Dynamics in action: Intentional behavior as a complex system*. Cambridge, MA: MIT Press.

- Kant, I. (1998). *The critique of judgement*. Trans. James C. Meredith. Oxford: Clarendon Press.
- Kelly, S. (2000). Grasping at straws: Motor intentionality and the cognitive science of skillful action. In *Heidegger, coping, and cognitive science: Essays in honor of Hubert L. Dreyfus*, vol. II, ed. M. Wrathall and J. Malpas, 161–177. Cambridge, MA: MIT Press.
- Kelso, J. A. S. (1995). *Dynamic patterns: The self-organization of brain and behavior*. Cambridge, MA: MIT Press.
- Kendon, A. (1990). *Conducting interaction: Patterns of behavior in focused encounters*. Cambridge: Cambridge University Press.
- Kim, J. (1999). Making sense of emergence. *Philosophical Studies* 95:3–36.
- Lakoff, G. (1987). *Women, fire, and dangerous things: What categories reveal about the mind*. Chicago: University of Chicago Press.
- Le Van Quyen, M., and Petitmengin, C. (2002). Neuronal dynamics and conscious experience: An example of reciprocal causation before epileptic seizures. *Phenomenology and the Cognitive Sciences* 1:169–180.
- Lenay, C. (2003). Ignorance et suppléance: La question de l'espace (Ignorance and augmentation: The question of space). HDR 2002, Université de Technologie de Compiègne.
- Leslie, A. (1987). Pretence and representations: The origins of "Theory of Mind." *Psychological Review* 94:412–426.
- Lewontin, R. C. (1983). The organism as the subject and object of evolution. *Scientia* 118:63–82.
- Lutz, A. (2002). Toward a neurophenomenology as an account of generative passages: A first empirical case study. *Phenomenology and the Cognitive Sciences* 1: 133–167.
- Maynard Smith, J., and Szathmáry, E. (1995). *The major transitions in evolution*. Oxford: W. H. Freeman.
- McGann, M., and De Jaegher, H. (2009). Self-other contingencies: Enacting social perception. *Phenomenology and the Cognitive Sciences* 8 (4): 417–437.
- Merleau-Ponty, M. (1962). *Phenomenology of perception*. London: Routledge.
- Miller, S. (1973). Ends, means, and galumphing: Some leitmotifs of play. *American Anthropologist* 75:87–98.
- Millikan, R. G. (1984). *Language, thought and other biological categories: New foundations for realism*. Cambridge, MA: MIT Press.

- Moran, G., Fentress, J. C., and Golani, I. (1981). A description of relational patterns during “ritualized fighting” in wolves. *Animal Behaviour* 29:1146–1165.
- Moreno, A., and Etxeberria, A. (2005). Agency in natural and artificial systems. *Artificial Life* 11:161–176.
- Nolfi, S., and Floreano, D. (2000). *Evolutionary robotics. The biology, intelligence, and technology of self-organizing machines*. Cambridge, MA: MIT Press.
- O’Regan, J. K., and Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences* 24 (5): 883–917.
- Oyama, S. (2000). *The ontogeny of information: Developmental systems and evolution*. 2nd edition. Durham, NC: Duke University Press.
- Pfeifer, R., and Scheier, C. (1999). *Understanding Intelligence*. Cambridge, MA: MIT Press.
- Piaget, J. (1936). *La naissance de l’intelligence chez l’enfant*. Neuchâtel-Paris: Delachaux et Niestlé.
- Piaget, J. (1951). *Play, dreams and imitation in childhood*. London: Routledge.
- Piaget, J. (1967). *Biologie et connaissance: Essai sur les relations entre les régulations organiques et les processus cognitifs*. Paris: Gallimard.
- Poincaré, H. (1907). *La science et l’hypothèse*. Paris: Flammarion.
- Rodriguez, E., George, N., Lachaux, J.-P., Matinerie, J., Reanault, B., and Varela, F. J. (2001). Perception’s shadow: long-distance synchronization of human brain activity. *Nature* 397:430–433.
- Rohde, M. (2010). *Enaction, embodiment, evolutionary robotics. Simulation models for a post-cognitivist science of mind*. Thinking Machines book series. Amsterdam and Paris: Atlantis Press.
- Rutkowska, J. C. (1997). What’s value worth? Constraining unsupervised behaviour acquisition. In *Fourth European Conference on Artificial Life*, ed. P. Husbands and I. Harvey, 290–298. Cambridge, MA: MIT Press.
- Sacks, H., Schegloff, E. A., and Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language* 50 (4): 696–735.
- Salomon, R. (1998). Achieving robust behavior by using proprioceptive activity patterns. *BioSystems* 47:193–206.
- Schiffrin, D. (1994). *Approaches to discourse*. Oxford: Blackwell.
- Schwartzman, H. B. (1978). *Transformations: The anthropology of children’s of play*. New York: Plenum.

- Sheets-Johnstone, M. (2003). Child's play: A multidisciplinary perspective. *Human Studies* 26:409–430.
- Silberstein, M., and McGeever, J. (1999). The search for ontological emergence. *Philosophical Quarterly* 49:182–200.
- Skarda, C. A., and Freeman, W. J. (1987). How brains make chaos in order to make sense of the world. *Behavioral and Brain Sciences* 10:161–195.
- Sporns, O., and Edelman, G. M. (1993). Solving Bernstein's problem: A proposal for the development of coordinated movement by selection. *Child Development* 64:960–981.
- Stewart, J., and Coutinho, A. (2004). The affirmation of self: A new perspective on the immune system. *Artificial Life* 10:261–267.
- Sutton-Smith, B. (1966). Piaget on play: a critique. *Psychological Review* 73:104–110.
- Sutton-Smith, B. (1997). *The ambiguity of play*. Cambridge, MA: Harvard University Press.
- Thelen, E., and Smith, L. B. (1994). *A dynamic systems approach to the development of cognition and action*. Cambridge, MA: MIT Press.
- Thompson, E. (2001). Empathy and consciousness. In *Between ourselves: Second-person issues in the study of consciousness*, ed. E. Thompson, 1–32. Exeter, UK: Imprint Academic.
- Thompson, E. (2005). Sensorimotor subjectivity and the enactive approach to experience. *Phenomenology and the Cognitive Sciences* 4:407–427.
- Thompson, E. (2007). *Mind in life: Biology, phenomenology, and the sciences of mind*. Cambridge, MA: Harvard University Press.
- Thompson, E., and Varela, F. (2001). Radical embodiment: Neural dynamics and consciousness. *Trends in Cognitive Sciences* 5:418–425.
- Trevarthen, C. (1979). Communication and cooperation in early infancy: A description of primary intersubjectivity. In *Before speech*, ed. M. Bullowa, 39–52. Cambridge: Cambridge University Press.
- van Gelder, T. (1999). Wooden iron? Husserlian phenomenology meets cognitive science. In *Naturalizing phenomenology: Issues in contemporary phenomenology and cognitive science*, ed. J. Petitot, F. J. Varela, B. Pachoud, and J.-M. Roy, 245–265. Stanford, CA: Stanford University Press.
- Varela, F. J. (1979). *Principles of biological autonomy*. New York: Elsevier North Holland.
- Varela, F. J. (1991). Organism: A meshwork of selfless selves. In *Organism and the origin of the self*, ed. A. I. Tauber, 79–107. Netherlands: Kluwer Academic.

Varela, F. J. (1996). Neurophenomenology: A methodological remedy for the hard problem. *Journal of Consciousness Studies* 3:330–350.

Varela, F. J. (1997). Patterns of life: Intertwining identity and cognition. *Brain and Cognition* 34:72–87.

Varela, F. J. (1999). The specious present: A neurophenomenology of time consciousness. In *Naturalizing phenomenology: Issues in contemporary phenomenology and cognitive science*, ed. J. Petitot, F. J. Varela, B. Pachoud, and J.-M. Roy, 266–314. Stanford, CA: Stanford University Press.

Varela, F. J., Lachaux, J.-P., Rodriguez, E., and Matinerie, J. (2001). The brainweb: Phase synchronization and large-scale integration. *Nature Reviews Neuroscience* 2:229–230.

Varela, F. J., Thompson, E., and Rosch, E. (1991). *The embodied mind: Cognitive science and human experience*. Cambridge, MA: MIT Press.

Verschure, P. J., Wray, J., Sporns, O., Tononi, G., and Edelman, G. M. (1995). Multilevel analysis of classical conditioning in a behaving real world artifact. *Robotics and Autonomous Systems* 16:247–265.

Vygotsky, L. S. (1966). Play and its role in the mental development of the child. *Social Psychology* 12:62–76.

Walter, W. G. (1950). An imitation of life. *Scientific American* 182 (5): 42–45.

Watson, M. W., and Jackowitz, E. R. (1984). Agents and recipient objects in the development of early symbolic play. *Child Development* 55:1091–1097.

Weber, A. (2003). *Natur als Bedeutung. Versuch einer Semiotischen Theorie des Lebendigen*. Würzburg: Königshausen and Neumann.

Weber, A., and Varela, F. J. (2002). Life after Kant: Natural purposes and the auto-poietic foundations of biological individuality. *Phenomenology and the Cognitive Sciences* 1:97–125.

Wheeler, M. (2005). *Reconstructing the cognitive world: The next step*. Cambridge, MA: MIT Press.

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review* 9:625–636.

Winnicott, D. (1971). *Playing and reality*. London: Routledge.

Winograd, T., and Flores, F. (1986). *Understanding computers and cognition*. Norwood, NJ: Ablex.