Dynamical Systems and Embedded Cognition

Randall D. Beer
Cognitive Science Program
Dept. of Computer Science
Dept. of Informatics
Indiana University
Bloomington, IN 47406

Draft – 4/18/07

To appear in K. Frankish and W. Ramsey (Eds.), *The Cambridge Handbook of Artificial Intelligence*. Cambridge University Press.

Please address all correspondence to:

Randall D. Beer
Cognitive Science Program
1910 E. 10th St. – 840 Eigenmann
Indiana University
Bloomington, IN 47406

Phone: (812) 856-0873
Fax: (812) 855-1086
Email: rdbeer@indiana.edu
URL: http://mypage.iu.edu/~rdbeer/

1. Introduction

The conceptual frameworks that we bring to our study of cognition can have a tremendous impact on the nature of that study. They provide a set of filters through which we view the world, influencing our choice of phenomena to study, the language in which we describe these phenomena, the questions we ask about them, and our interpretations of the answers we receive. For much of the last fifty years, thinking about thinking has been dominated by the computational framework, the idea that systems are intelligent to the extent that they can encode knowledge in symbolic representations which are then algorithmically manipulated so as to produce solutions to the problems that these systems encounter. More recently, the connectionist framework forced an important refinement of the computational framework in which representation and computation could be distributed across a large number of loosely neuron-like units.

Beginning around the mid 1980s, just as the popularity of connectionism was rising, another conceptual framework appeared (or, as in the case of connectionism, reappeared) on the scene. This framework, which, for want of a catchier label, I will call the situated, embodied, dynamical (SED) framework, focuses on concrete action and emphasizes the way in which an agent's behavior arises from the dynamical interaction between its brain, its body and its environment. In this chapter, I will attempt to trace some of the history of the individual intellectual threads of situated activity, embodiment and dynamics that underlie the SED approach. I will particularly focus on the years 1985-1995. Although there were important precursors to the SED approach (some of which I will briefly mention), and work in this area has grown rapidly in recent years, many of the pivotal ideas were first given their modern form during this decade.

2. Situated Activity

The first intellectual thread comprising the SED approach is *situated activity*. Roughly speaking, situated activity stresses three ideas that have been traditionally neglected in AI and cognitive science.

S1) <u>Concrete action</u>. Actually taking action in the world is more fundamental than the abstract descriptions that we sometimes make of it. While conscious deliberation clearly has its role, the ultimate job of an intelligent agent is to *do* something, to take some concrete action with consequences beyond its own skull.

S2) <u>Situatedness</u>. An agent's immediate environment plays a central role in its behavior. This environment is not only a rich source of constraints and opportunities for the agent, but also a context that gives meaning to the agent's actions.

S3) <u>Interactionism</u>. An agent's relationship with its environment is one of ongoing interaction. The environment does not serve merely as a source of isolated problems for the agent to solve, but rather a partner with which the agent is fully engaged in moment-to-moment improvisation.

The philosophical roots of situated activity can be traced to phenomenology, especially the work of Heidegger (1927/1962), which was brought into AI and cognitive science primarily through the criticisms of Dreyfus (1972/1992). One of Heidegger's key insights was the distinction he drew between objects being *zuhanden* ("ready-to-hand") and *vorhanden* ("present-at-hand"). In our normal daily experience, we usually encounter things as resources for immediate action in the service of achieving our goals. For example, to someone in the act of hammering a nail, the hammer in some sense ceases to exist. Rather, like any tool, it becomes merely an extension of the arm (i.e., it is ready-to-hand). It is only when we explicitly adopt an intellectual attitude toward the hammer (e.g., because the handle has broken and the hammer is suddenly unable to perform its normal function), that the hammer emerges from the unarticulated background of things as a distinct object characterized by its own set of properties (i.e., it becomes present-at-hand). A number of authors have carefully articulated the challenges that phenomenological ideas pose to the objectivist, Cartesian worldview that has dominated thinking in AI and cognitive science (Winograd and Flores, 1986; Dreyfus, 1972/1992; Varela, Thompson and Rosch, 1991; Clark, 1997; Wheeler, 2005).

Another important precursor to situated activity was Gibson's Ecological Psychology (Gibson, 1979). Based on his studies of vision in World War II pilots, Gibson emphasized the structure inherent in an organism's environment and the importance of the organism/environment relation to a theory of perception. For example, the optic flow field of a visual animal moving through its environment carries a great deal of information about the direction and speed of motion, distances to objects, surfaces, textures, etc. Gibson's views eventually encompassed a wide-ranging rejection of cognitivism. However, for our purposes here, Gibson's most important contribution is his notion of affordances, the possibilities for action that an environment presents to an agent. For example, Heideggar's hammer affords pounding nails due to the graspability of its handle, and the shape and hardness of its head. Furthermore, Gibson argued that, although affordances are perceivable facts about the world, they are ecological in the sense that their significance is relative to the capabilities of a particular organism. For example, an opening that affords passability to a mouse does not necessarily afford passability to a human being.

A third important influence on situated activity came from work in the social sciences, especially ethnomethodological approaches. For example, Suchman, an anthropologist studying man-machine interaction, traced breakdowns in communication between a person and a help system for a photocopy machine to mistaken assumptions made by the designers of the system about the nature of action (Suchman, 1987). She rejected the traditional view in AI and cognitive science that action results from the execution of a plan and argued instead that action must be understood as situated in the sense that it is contingent upon the actual circumstances as they unfold. On this view, explicit plans are best interpreted as resources for communicating about action rather than as mechanisms for action. Based on his studies of the navigation team of a large naval vessel, another anthropologist, Hutchins, similarly concluded that cognition “in the wild” must be understood as a culturally constituted activity depending heavily on the unfolding situation in which it occurs (Hutchins, 1995).

Within AI, situated ideas came to the fore in the mid 80s. Earlier demonstrations of how rich behavior could arise from simple mechanisms interacting with complex environments include W. Grey Walter’s robotic “tortoises” (Walter, 1953) and Braitenburg’s simple “vehicles” (Braitenburg, 1984). Another important precursor was work on situation semantics (Barwise and Perry, 1983). However, situated activity research within AI arose mainly as a reaction against the traditional planning view of action. Agre and Chapman stressed the inability of classical planning techniques to scale to complex, uncertain, real-time environments and proposed instead that routine activity arises from the interaction of simple internal machinery with the immediate situation (Agre and Chapman, 1987; Agre, 1997). Agre and Chapman demonstrated the utility of this idea in a series of programs, the most well-know of which was Pengi, an agent that played the video arcade game Pengo in real-time despite having to deal with hundreds of often unpredictable objects. Rosenschein and Kaelbling showed how a specification of an agent’s goals could be “compiled away” into simple machinery such that, although it still made sense for an external observer to talk about the agent's knowledge and beliefs, they no longer played any direct role in the agent's actions (Rosenschein and Kaelbling, 1986). Brooks' influential work on autonomous robots rejected the traditional sense-model-plan-act cycle, emphasizing that often "the world is its own best model" (Brooks, 1986; 1991a). He developed a layered control system known as the subsumption architecture, in which networks of finite state machines augmented by timers and register interact with one another and the immediate circumstances to produce behavior, and deployed it on a variety of different robots. Cliff (Cliff, 1991) and I (Beer, 1990) demonstrated the significant potential for interaction between work on the neural basis of animal behavior and situated agents, developing computational neuroethology models of a hoverfly and a cockroach, respectively.

No one would presumably deny that the environmental situation has an important role to play in an agent's behavior, but just how fundamental this observation is remains controversial (Kirsch, 1991; Vera and Simon, 1993; Hayes, Ford and Agnew, 1994; Clancey, 1997; Anderson, 2003). To some, situated activity smacks of behaviorism, but this charge depends a great deal on what exactly one means by behaviorism. It is certainly true that work in situated activity exhibits a renewed emphasis on concrete behavior over abstract reasoning. However, abstract reasoning is not rejected by situated approaches, but rather relegated to a supporting role as an evolutionarily recent elaboration of a more basic capacity for getting around in the world. It is also true that much work in situated activity has tended to emphasize reactive architectures, in which an agent's actions are completely determined by its sensations, and to either reject or at least significantly reconstrue the idea of internal representations. Reactive architectures are strongly reminiscent of the stimulus-response paradigm embraced by behaviorism, and have well-known limitations when it comes to, for example, anticipatory behavior. However, as we shall see later in this chapter, a commitment to purely reactive architectures is unnecessary, and it is possible to articulate a role for internal state that is both essential and interestingly different from the representational role that such state plays in traditional AI and cognitive science.

Perhaps the most controversial idea that has emerged from research on situated cognition in recent years is the notion of the extended mind (Clark, 1997; Clark & Chalmers, 1998). The basic idea here is that not only does an agent's environment play an essential role in its behavior, but the agent itself can manipulate that role by actively organizing its environment so as to increase its problem-solving ability. For example, we lay out the ingredients for a recipe in the order in which they will be needed, and we use maps to find our way through sprawling cities. Such scaffolding allows us to offload significant parts of our cognitive processing into the environment. Furthermore, through language, we can coordinate the activities of many people so that they can collectively accomplish things that no individual person may be able to, such as navigating a large naval vessel (Hutchins, 1995). Extended mind advocates argue that if memory, problem-solving, etc. can be spread across many agents and artifacts, then cognition itself must be understood as a distributed phenomenon that transcends the skull of an individual agent, and properly belongs only to the larger system of agents and artifacts of which that individual is a part. Indeed, even social insects are known to collectively accomplish complex construction tasks such as nest-building by modifying their environment in such a way as to appropriately organize the flow of workers and material, a process referred to as stigmergy (Turner, 2000).

3. Embodiment

A second intellectual thread comprising the SED approach is *embodiment*. There are at least three somewhat distinct ideas that have been advanced by advocates of embodied cognitive science.

E1) Physical embodiment. The uniquely physical aspects of an agent's body are crucial to its behavior, including its material properties, the capabilities for action provided by the layout and characteristics of its degrees of freedom and actuators, the unique perspective provided by the particular layout and characteristics of its sensors, and the modes of sensorimotor interaction that the sensors and actuators collectively support.

E2) Biological embodiment. Not only are the physical characteristics of bodies important, but the specifically biological facts of an organism's existence must also be taken into account, including the relevant neuroscience, physiology, development and evolution.

E3) Conceptual embodiment. Even when engaged in pure ratiocination, our most abstract concepts are still ultimately grounded in our bodily experiences and body-oriented metaphors.

The philosophical roots of embodiment can also be traced to phenomenology, especially the work of Merleau-Ponty (1962), who made bodily involvement in the world central to his phenomenology of lived experience. To take but one example, Merleau-Ponty's argument that how we perceive an object is shaped by the kinds of interactions with it that our body allows can be seen as an early precursor to Gibson's (1979) notion of affordances. Merleau-Ponty's thought also played a major role in Dreyfus' critique of computational theories of mind (Dreyfus, 1972/1992).

Within AI and cognitive science, the importance of physical embodiment was first emphasized by Brooks (1991b). Brooks argued that AI needed to move beyond the abstract microworlds that had been its primary concern and begin to address the sorts of problems encountered by real robots moving around in real environments. In this way, Brooks suggested, the extent to which most classical AI techniques are simply untenable in realistic situations would become clear. In its milder form, the argument of physical embodiment is simply that the material properties of the body and environment play a key role in its behavior and, by building robots, we get this physics "for free" rather than having to painstakingly model it. In its most radical form, the claim is that only physically instantiated AI systems will exhibit truly intelligent behavior. Coupled with the contemporaneous trends in situated cognition reviewed in the previous section, Brooks' arguments unleashed an explosion of work in behavior-based robotics (Arkin, 1998), active perception (Ballard, 1991; Churchland, Ramachandran and Sejnowski, 1994; Noë, 2004), embodied cognitive science (Pfeifer and Scheier, 1999), autonomous agents (Maes, 1990), some aspects of artificial life (Langton, 1989) and the philosophy of mind (Clark, 1997).

Biological embodiment takes the arguments of physical embodiment one step further. Not only are the physical characteristics of bodies important, but so are the biological facts of an organism's existence. The conditions necessary to maintain our living state fundamentally constrain our behavioral and cognitive capacities. In addition, the specific properties of bone, muscle and skin, the specific characteristics of biological sensors, and the ways these sensory and motor capabilities are knitted together in human bodies fundamentally define our own particular mode of embodiment. Furthermore, the fact that we have gone through the particular evolutionary and developmental history that we have may also have important consequences for our behavioral and cognitive architecture. For example, Thelen and Smith have argued for the importance of understanding the sensorimotor origins of cognition in development, both in studies of the development of walking in infants (Thelen and Smith, 1994) and, more recently, in studies of Piaget's classic A-not-B error (Thelen, Schöner, Scheier and Smith, 2001). A similar argument can be made for the emergence in evolution of uniquely human cognitive capacities from simpler precursors. Finally, there has been a very strong push toward incorporating more neurobiological realism into embodied agents (Arbib, 1987; Beer, 1990; Edelman et al., 1992; Webb, 2001). Conversely, neuroscience has begun to take seriously the role of the body and of neuromechanical interactions in the production of behavior (Chiel and Beer, 1997).

Thus, the conventional claim of biological embodiment is that the biological features of organisms matter to their behavior and cognition. A more radical claim that is sometimes associated with biological embodiment is that the living state itself is fundamental to cognition (Maturana and Varele, 1980; Varela, Thompson and Rosch, 1991; Stewart, 1996; Moreno, Umerez and Ibañez, 1997; Di Paolo, 2005). The idea here is generally not that the material or biochemistry of life is essential, but rather that the organization of living systems is indispensable to their cognitive capabilities. The relevant notion of living organization is generally derived from Maturana and Varela's concept of autopoiesis (roughly, a self-producing network of components and processes, i.e., a kind of organizational homeostasis) (Maturana and Varela, 1980). The key idea of the strong biological embodiment claim is that, in order to be truly autonomous, a system must be capable of creating its own laws, rather than simply having them imposed by an external observer or designer.

Finally, conceptual embodiment concerns the way in which even abstract concepts are often grounded in bodily experience and metaphor. For example, Harnad defined the symbol grounding problem as the problem of how words, and ultimately mental states, get their meaning (Harnad, 1990), and he proposed that a way to address this problem is to ground them in sensorimotor signals. Futhermore, Lakoff and Johnson have argued that the structure of our reason is grounded in the details of our embodiment, and that many abstract concepts are metaphors derived from sensorimotor domains (Lakoff and Johnson, 1999). For example, we speak of understanding something as “grasping” it and we speak of failing to understand something as a failure to “grasp” it or it “going over our heads”. Likewise, bad things “stink” and the “pieces” of a theory “fit” together.

4. Dynamics

The final intellectual thread comprising the SED approach is *dynamics*, within which we must distinguish at least three ideas.

D1) *Dynamical systems theory* (DST). A mathematical theory that can be applied to any system characterized by a state that changes over time in some systematic way.

D2) *The dynamical framework*. A collection of concepts, intuitions, and metaphors involved in taking a dynamical perspective on some system of interest.

D3) *The dynamical hypothesis*. A specific hypothesis, put forward by Tim van Gelder (1998), for how DST and the dynamical framework could be combined into a rigorous counterproposal to the traditional computational hypothesis in AI and cognitive science.

A dynamical system is a mathematical abstraction that unambiguously describes how the state of some system evolves over time (Abraham and Shaw, 1992; Strogatz, 1994; Kuznetsov, 2004). It consists of a state space S, an ordered time set T, and an evolution operator ϕ that transforms a state at one time to another state at some other time. A dynamical system whose evolution depends only its internal state is called autonomous, while one whose evolution also depends on external inputs is called nonautonomous. S can be numerical or symbolic, continuous or discrete (or a hybrid of the two) and of any topology and dimension (including infinite dimensional). T is typically either the set of integers or the set of real numbers. The evolution operator may be given explicitly or defined implicitly and it may be deterministic or stochastic.

The most common examples of dynamical systems are sets of ordinary differential equations and iterated maps, but many other kinds of mathematical systems, such as finite state machines, cellular automata, Turing machines, and sets of stochastic or partial differential equations can also be fruitfully described and analyzed in dynamical terms. For any mathematical system that can be put into this form, DST offers a wide variety of tools for analyzing the temporal behavior of such systems, many of which were first developed by the French mathematician Henri Poincaré in support of his work in celestial mechanics. These tools include the identification of invariant sets (e.g., fixed points, limit cycles, etc.), and the characterization of their local (e.g., stability) and global (e.g., saddle manifolds) structure, and their dependence on parameters (e.g., bifurcations). It is important to reiterate that, just like the formal theory of computation, DST is a body of mathematics, and not a scientific theory of the natural world.

Despite the fact that DST is not itself a scientific theory, taking a dynamical perspective brings with it a set of concepts, intuitions and metaphors – a certain worldview – that influences the questions we ask, the analyses we perform and how we interpret the results (van Gelder, 1995). When one approaches some system from a computational perspective, one is concerned with what function the system is trying to compute, in what format the problem input is specified, in what output format the answer is required, how the relevant features of the problem are to be represented, by what algorithms these representations are to be transformed, and how the performance of these algorithms scales with problem size. In contrast, when one approaches some system from a dynamical perspective, one seeks to identify a minimal set of state variables whose evolution can account for the observed behavior, the dynamical laws by which the values of these variables evolve in time, the overall spatiotemporal structure of their possible evolution, and the sensitivity of this structure to variations in inputs, states and parameters.

The dynamical perspective has been found to be a fruitful one in many areas of cognitive science (Port and van Gelder, 1995; Beer, 2000). A dynamical perspective on brain and behavior was first explicitly articulated by Ashby (Ashby, 1960). Within neural networks, Grossberg has long emphasized the importance of dynamical ideas (Grossberg, 1969). Indeed, DST is now an essential tool in computational neuroscience (Izhikevich, 2007), not just for analyzing individual nerve cells or small circuits, but also entire brain systems (Skarda and Freeman, 1987). Dynamical ideas were first brought into ecological psychology by Kugler (Kugler, Kelso and Turvey, 1980; for reviews see Turvey, 1990 and Warren, 2006). Kelso and colleagues have pursued a dynamical perspective on brain and behavior for many years, especially emphasizing the role of self-organization in the creation of behavioral patterns and the transitions between them (Kelso, 1995; Schöner and Kelso, 1988). Thelen and Smith have argued for a dynamical approach to cognitive development, in which processes and change are studied using the same tools across a range of timescales (Thelen and Smith, 1994). Elman emphasized the fundamentally temporal character of language understanding, with preceding words strongly influencing the interpretation of subsequent ones, and has developed a dynamical approach to language (Elman, 1995). Finally, I argued that dynamical systems theory provides the appropriate theoretical language and tools for analyzing the kinds of autonomous agents that were being developed in AI and robotics (Beer, 1992; 1995a), and Smithers (1995) and Schöner (1995) advocated a dynamical approach to the design of autonomous robots.

A specific formulation that has received a great deal of attention is the dynamical hypothesis put forward by van Gelder (van Gelder, 1995; 1998). Van Gelder defines a dynamical system as a quantitative system, that is, a system whose state space, time set and evolution law involve numerical quantities. As we saw above, this is a significant restriction of the mathematical definition of a dynamical system. His dynamical hypothesis then has two components: (1) the nature hypothesis and (2) the knowledge hypothesis. The claim of the nature hypothesis is ontological: cognitive systems are dynamical systems. In contrast, the knowledge hypothesis claims only that cognitive systems are best understood using the tools of dynamical systems theory. Given that even many advocates of the dynamical approach do not fully support van Gelder's dynamical hypothesis, it is unfortunate that most critical discussion of the dynamical approach to cognition has focused on van Gelder's specific formulation (Grush, 1997; Eliasmith, 1997; Bechtel, 1998; Van Leeuwen, 2005). Nevertheless, it is an historically important attempt to formulate a dynamical alternative to the computational hypothesis.

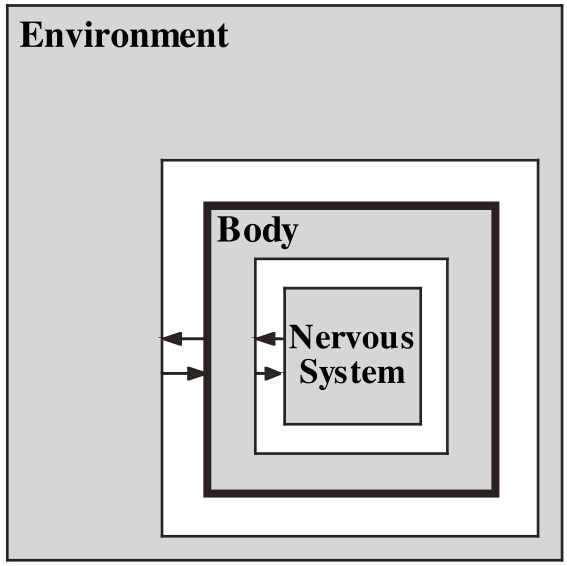

Figure 1: An agent and its environment are coupled dynamical systems. The agent in turn is composed of coupled nervous system and body dynamical systems.

5. Toward An Integrated Perspective

To this point, I have treated situatedness, embodiment and dynamics as relatively separate intellectual threads. I did this both because the historical development of these ideas occurred somewhat independently and because they are logically independent, that is, people can and do hold each of them individually without necessarily also subscribing to the others. However, it will not have escaped the careful reader's attention that there is a great deal of potential overlap and synergism between them. The goal of this section is to articulate an integrated theoretical framework that combines the insights from situatedness, embodiment and dynamics. In contrast to previous sections, I will also adopt a more personal viewpoint in this section, describing my own particular integrative view (Beer, 1992; 1995a; 1995b; 1997; 2003) rather than attempting a general survey of all such views.

The basic situated, embodied, dynamical (SED) framework is quite simple and is illustrated in Figure 1. It consists of the following three postulates:

SED1) *Brains, bodies and environments are dynamical systems* (cf. S2, E1, E2, D1, D2). Nervous systems, bodies and environments are all conceptualized as dynamical systems, by which I mean only that we assume that each can be characterized by a set of states whose temporal evolution is governed by dynamical laws.

SED2) *Brain, body and environment dynamics are coupled* (cf. S1, S3, D1, D2). Nervous systems are embodied in bodies, which are in turn situated within environments, leading to dense interaction between these three component systems. The coupled brain-body subsystem will be termed the "agent". Coupling that flows from the environment to the agent will be termed "sensory" and coupling that flows in the opposite direction will be termed "motor". The "behavior" of an agent will be defined as its trajectory of motor actions.

SED3) *The agent is subject to viability constraints* (cf. E2). There are conditions on the dynamics of the agent that determine its viability. If these viability constraints are violated, then the agent ceases to exist as an independent entity and can no longer engage in behavioral interactions with its environment. We will not consider this postulate further here, but discussion of its role in the SED framework can be found in (Beer, 2004).

The a priori theoretical commitments of this framework are quite minimal. Indeed, it is hard to imagine a theoretical framework that makes fewer commitments than this. What could possibly follow from such a small set of claims? In fact, quite a number of nontrivial consequences follow almost immediately if we take these three postulates seriously.

Perhaps the most important conclusion is this: Strictly speaking, *behavior is a property of the entire coupled brain-body-environment system*, and cannot in general be properly attributed to any one subsystem in isolation from the others. We have defined behavior to be only the trajectory of an agent's motor actions. However, because the brain, body and environment dynamics are coupled, they form a single larger autonomous dynamical system with its own trajectories of temporal evolution. The trajectories of an agent's motor actions are merely projections of the full trajectories of the complete brain-body-environment system, and it is these full trajectories that are the proper objects of study within the SED framework.

Even though behavior is a property of the entire coupled system, it is still meaningful to ask about the relative contributions of brain, body and environment to some particular feature of a behavioral trajectory. In order to do so, we must open the coupled brain-body-environment system by cutting one or more of the coupling pathways in order to isolate the component we wish to study. This component then becomes a nonautonomous dynamical system and our analysis involves examining how its own intrinsic dynamics interacts with the inputs it receives from the other components of the coupled system in the production of the behavioral feature of interest. This has many interesting consequences for the way we conceive of traditional behavioral and cognitive phenomena.

For example, perception is generally viewed as a means by which an agent extracts from the raw sensory signals and internally represents the structure of its environment. But a dynamical systems follows a trajectory specified by its own internal state and dynamical laws. Sensory inputs cannot in general place a nonautonomous dynamical system into some state uniquely characteristic of a given external object. Rather, the most that they can do is bias the intrinsic tendencies of the agent dynamics by selecting some particular trajectory from the set of possible trajectories that the agent's dynamical laws allow from its current state. This suggests a more behavior-oriented view of perception that is reminiscent of Gibson (1979). On this view, perception is a process whereby agent dynamics that is appropriately sensitive to environmental influences is perturbed by the trajectory of sensory inputs that it receives into behavior appropriate to its circumstances. Furthermore, because the coupling between an agent and its environment is two-way, an agent's action can shape its own perception. Agents not only perceive in order to act, but they also act in order to perceive.

Because agents in the SED framework are dynamical, they are not vulnerable to the criticisms that have been leveled against reactive agents. A reactive agent is one whose motor outputs depend only on its sensory inputs; it is merely a function from sensation to action. Although such an agent can participate in complex interactions when coupled to a dynamic environment, its behavior is always subordinated to that environment since it possesses no dynamics of its own. In contrast, the response of a dynamical agent is determined at least in part by its own internal dynamics. Because it possesses internal state, a dynamical agent can respond differently to the same sensory stimulus at different times, it can initiate behavior independently of its immediate environment, it can modify its behavior based on its history of interactions, and it can exploit long-term correlations in its environment to organize its behavior in anticipation of future events.

One significant advantage of the SED framework is that it offers the possibility of a uniform treatment of disparate behavioral and cognitive phenomena that have often been seen as irreconcilable. At one extreme, some basic sensorimotor behavior may be mostly reactive in character, with internal state playing only a small role in "coloring" the agent's responses to its environment. At the other extreme, some of our most cognitive behavior can be conceived as being nearly decoupled from the immediate environment circumstances, driven primarily by the temporal evolution of internal state. Of course, most behavior is usually a mixture of external and internal influences, with the relative importance of the two varying, sometimes substantially, from moment to moment. Indeed, the interesting questions of how higher cognitive processes arose from more basic sensorimotor competence during the course of evolution and development seems much more approachable within a theoretical framework that places them both on a common footing. On this view, higher cognition does not necessarily alter our fundamentally situated, embodied and dynamic character, but instead augments it with a vastly increased reservoir of internal dynamics.

How are we to understand the nature and role of this internal state within a dynamical agent? The traditional computational interpretation of such states would be as internal representations. But internal state is a property of physical systems in general, and these states can covary with states outside the system in quite complicated ways. Unless we wish to grant representational status to all physical states (does a thunderstorm represent the topography of the terrain over which it passes?), there must be additional conditions that license the modifier "representational". Unfortunately, despite the fundamental role that the notion of representation plays in computational approaches, there is very little agreement about what those additional conditions might be. These considerations have led me to adopt a position of representational skepticism (not, as some have suggested, antirepresentationalism) (Beer, 2003). I view the representational status of an internal state as an empirical question, to be settled according to the precise definition of the particular representational notion on offer. Thus, by not taking representation for granted, a dynamical perspective offers a broader theoretical playing field. On the one hand, it offers the possibility of understanding what representations are and when and how they arise. On the other hand, we may find that, at least in some cases, the roles played by the internal states of a dynamical agent simply cannot be usefully interpreted as representational.

What is the relationship between a SED approach to cognition and the more familiar computational and connectionist approaches? Such a comparison is fraught with difficulties. For example, we must distinguish between the bodies of mathematics that underlie each of these approaches and the theoretical claims that these approaches make. Mathematically, these systems are all of roughly equivalent power. All finite state machines and Turing machines, as well as all recurrent connectionist networks, are dynamical systems. All dynamical systems can be approximated by traditional Turing machines, and Turing machines defined over the reals are equivalent to dynamical systems. Feedforward connectionist networks can approximate arbitrary functions, which can be used to implement the control logic of a Turing machine and can approximate arbitrary dynamical systems when the output of the network is fed back into its input. Recurrent neural networks are known to be universal dynamics approximators. And so on. Thus, there is no useful mathematical distinction to be drawn between these different approaches. This, I think, is one of the ways in which van Gelder's dynamical hypothesis goes wrong (Beer, 1998).

In addition, we must recognize that computationalism, connectionism and dynamicism are not really scientific theories at all, because they themselves do not make falsifiable predictions. Rather, they are what I have called theoretical frameworks (Beer, 1995b). They provide a set of pretheoretical intuitions, a theoretical vocabulary, a style of explanation, a worldview within which particular falsifiable theories of specific cognitive phenomena are formulated and analyzed. The computationalist framework, for example, emphasizes the structure and content of the internal representations used by an agent and the algorithms by which those representations are manipulated. In contrast, the connectionist framework emphasizes the network architecture, the learning algorithm, the training protocol and the intermediate distributed representations that are developed. In this sense, many connectionist models are still disembodied, unsituated, and computational (albeit distributed) in nature (Harvey, 1992/1996). Finally, the SED framework emphasizes the structure of the space of all possible trajectories of the brain-body-environment system and the various forces, both internal and external to the agent, that shape those trajectories so as to stabilize some particular pattern of behavior. It is likely that all three perspectives will be important in any future theory of behavior and cognition. For example, since the neural components of a SED model are often recurrent connectionist networks and deliberative reasoning is one of the cognitive phenomena that must eventually be addressed, ideas and mathematical tools from both computationalism and connectionism are likely to play an essential role even in a SED-centered theory. The exact mix of insights from these three theoretical frameworks (or other frameworks yet unimagined!) that will ultimately prove to be the most fruitful remains an open question that only ongoing empirical investigation can resolve.

6. Methodological Issues

Taking the SED framework seriously raises many difficult methodological issues. Studying just one component of a brain-body-environment system is difficult enough, but studying the interactions of all three simultaneously is a daunting task. Experimentally, we currently lack the instruments to monitor and manipulate the activity of all the relevant neurons within the nervous systems of intact, behaving animals, let alone the relevant properties of the animal's body and environment. Theoretically, we currently lack the mathematical tools necessary to understand large networks of densely-interconnected, heterogeneous, nonlinear dynamical elements, particularly in systems that were evolved for their behavioral efficacy and not for their intelligibility in terms of traditional engineering design principles of modularity and hierarchical decomposition.

For these reasons, a number of researchers have turned to the study of model agents using dynamical neural networks and evolutionary algorithms (Beer and Gallagher, 1992; Cliff, Harvey and Husbands, 1993; Nolfi and Floreano, 2000, Beer, 2003; Harvey et al., 2005). In this approach, a model "nervous system" is embodied in a model body, which is in turn situated in a model environment. The entire system is evolved to perform some behavior of interest. A common choice of nervous system model is continuous-time recurrent neural networks (Beer, 1995c), which are known to be universal approximators of smooth dynamics (Kimura and Nakano, 1998). Typically, only the neural parameters are evolved, but in some work, network architecture and body properties are also evolved. One significant advantage of an evolutionary approach is that it minimizes a priori theoretical assumptions and thus allows the space of possible brain-body-environment systems capable of generating a particular behavior to be explored.

This evolutionary methodology has already been applied successfully to a wide range of interesting behavior (Nolfi and Floreano, 2000). A great deal of work has focused on sensorimotor behavior, such as chemotaxis, legged locomotion, object avoidance and navigation (Beer and Gallagher, 1992; Gruau, 1995; Kodjabachian and Meyer, 1998; Vaughan, Di Paolo and Harvey, 2004; Vickerstaff and Di Paolo, 2005). Another line of work has focused on the evolution of learning behavior (Miller and Todd, 1991; Yamauchi and Beer, 1994; Floreano and Mondada, 1996; Tuci, Quinn and Harvey, 2003; Izquierdo-Torres and Harvey, 2006). In addition, there has been considerable work on visually-guided behavior (Cliff, Harvey and Husbands, 1993) and its application to categorical perception, selective attention, and other cognitively-interesting tasks (Beer, 1996; Slocum, Downey and Beer, 2000; Beer, 2003; Di Paolo and Harvey, 2003; Ward and Ward, 2006). Finally, the evolution of communication has also been an active area of research (Werner and Dyer, 1992; Di Paolo, 1998; Di Paolo, 2000; Marocco, Cangleso & Nolfi, 2003; Steels, 2003; Nolfi, 2005). Thus, although there are difficult open issues in scaling evolutionary approaches to increasingly complicated behavior, one could argue that the agents that have already been evolved are interesting enough that their careful analysis could teach us many things about the dynamics of brain-body-environment systems.

Indeed, for me, the main interest is not in evolving such model agents per se, but rather in analyzing the resulting brain-body-environment systems using the tools of dynamical systems theory (Beer, 1995a,b,c; Husbands, Harvey and Cliff, 1995; Chiel, Beer and Gallagher, 1999; Beer, Chiel and Gallagher, 1999; Beer, 2003). The primary purpose of such an analysis is to build the intuitions, theoretical concepts and mathematical and computational tools necessary for understanding the dynamics of brain-body-environment systems. While DST provides a solid foundation for such investigations, many additional issues must be addressed. For example, there are different levels at which the dynamics of a brain-body-environment system can be analyzed, including the autonomous dynamics of the entire coupled system, how the coupled behavior arises from the interaction between the nonautonomous environment and agent dynamics, how the nonautonomous agent dynamics arises from the interaction between the nonautonomous body and neural dynamics, and how the nonautonomous neural dynamics arises from the architecture, intrinsic and synaptic parameters of the neural elements.

A final issue that must be addressed is understanding nonautonomous dynamics. The mathematical tools of DST are most highly developed in the case of autonomous dynamical systems, when the analysis can focus on attractors and their bifurcations. However, as mentioned above, when we wish to understand the contribution of a particular component of a brain-body-environment system, we must decompose the coupled system into interacting nonautonomous subsystems, and study their transient responses to time-varying inputs received from the other components. Unfortunately, the mathematical tools for analyzing transient dynamics require significant further development.

7. Prospects

Like both computationalism and connectionism, the situated, embodied and dynamical framework described in this chapter has its roots in ideas first articulated in the 1940s and 1950s. However, because the modern form of the SED framework only emerged in the years 1985-1995, it has had far less time for development than have the computational and connectionist frameworks. The number of people working within the SED framework is also considerably smaller at present. Despite these disadvantages, situated, embodied and dynamical ideas are having a major impact on thinking in cognitive science, AI and robotics, neuroscience, developmental psychology and philosophy of mind.

In order to further explore the scope and limits of the SED framework, and to clarify the best mix of computational, connectionist, and SED ideas necessary for understanding the mechanisms of behavior and cognition, considerable further development is necessary. First and foremost, this will require the construction and analysis of many more concrete model agents, especially those of a more cognitively-interesting nature. This in turn will require the continued development of techniques for scaling evolutionary techniques and dynamical analysis to larger systems and the further development of techniques for analyzing the transient dynamics of nonautonomous dynamical systems. Finally, there is a need for improved education in dynamical systems concepts within the cognitive science community, and for software to support the dynamical analysis of brain-body-environment systems.

References

Abraham, R.H. and Shaw, C.D. (1992). *Dynamics: The Geometry of Behavior*. Addison-Wesley.

Agre, P.E. (1997). *Computation and Human Experience*. Cambridge University Press.

Agre, P.E. and Chapman, D. (1987). Pengi: An implementation of a theory of activity. In *Proceedings of the Sixth National Conference on AI* (pp. 268-272). Morgan Kaufmann.

Anderson, M.L. (2003). Embodied cognition: A field guide. *Artificial Intelligence* **149**:91-130.

Arbib, M.A. (1987). Levels of modeling of mechanisms of visually-guided behavior. *Behav. Brain Sci*. **10**:407-465.

Arkin, R.C. (1998). *Behavior-Based Robotics*. MIT Press.

Ashby, W.R. (1960). *Design for a Brain* (2nd Edition). John W. Wiley & Sons.

Ballard, D.H. (1991). Animate vision. *Artificial Intelligence* **48**:57-86.

Barwise, J. and Perry, J. (1983). *Situations and attitudes*. MIT Press.

Bechtel, W. (1998). Representations and cognitive explanations: Assessing the dynamicist's challenge in cognitive science. *Cognitive Science* **22**:295-318.

Beer, R.D. (1990). *Intelligence as Adaptive Behavior: An Experiment in Computational Neuroethology*. Academic Press.

Beer, R.D. (1992). A dynamical systems perspective on autonomous agents. Technical Report CES-92-11, Dept. of Computer Engineering and Science, Case Western Reserve University.

Beer, R.D. (1995a). A dynamical systems perspective on agent-environment interaction. *Artificial Intelligence* **72**:173-215.

Beer, R.D,. (1995b). Computational and dynamical languages for autonomous agents. In R. Port and T. van Gelder (Eds.) (pp. 121-147).

Beer, R.D. (1995c). On the dynamics of small continuous-time recurrent neural networks. *Adaptive Behavior* **3**:471-511.

Beer, R.D. (1996). Toward the evolution of dynamical neural networks for minimally cognitive behavior. In P. Maes et al. (Eds.). *From Animals to Animats 4: Proceedings of the Fourth International Conference on Adaptive Behavior* (pp. 421-429). MIT Press.

Beer, R.D. (1997). The dynamics of adaptive behavior: A research program. *Robotics and Autonomous Systems* **20**:257-289.

Beer, R.D. (1998). Framing the debate between computational and dynamical approaches to cognitive science. *Behav. Brain Sci.* **21**:630.

Beer, R.D. (2000). Dynamical approaches to cognitive science. *Trends Cog. Sci.* **4**:91-99.

Beer, R.D. (2003). The dynamics of active categorical perception in an evolved model agent (with commentary and response). *Adaptive Behavior* **11**:209-243.

Beer, R.D. (2004). Autopoiesis and cognition in the game of Life. *Artificial Life* **10**:309-326.

Beer, R.D., Chiel, H.J. and Gallagher, J.C. (1999). Evolution and analysis of model CPGs for walking II. General principles and individual variability. *J. Computational Neuroscience* **7**:119-147.

Beer, R.D. and Gallagher, J.C. (1992). Evolving dynamical neural networks for adaptive behavior. *Adaptive Behavior* **1**:91-122.

Braitenburg, V. (1984). *Vehicles: Experiments in Synthetic Psychology*. MIT Press.

Brooks, R.A. (1986). A robust layered control system for a mobile robot. *IEEE J. Robot. Autom.* **2**:14-23.

Brooks, R.A. (1991a). Intelligence without representation. *Artificial Intelligence* **47**:139-159.

Brooks, R.A. (1991b). New approaches to robotics. *Science* **253**:1227-1232.

Chiel, H.J. and Beer, R.D. (1997). The brain has a body: Adaptive behavior emerges from interactions of nervous system, body and environment. *Trends Neurosci.* **20**:553-557.

Chiel, H.J., Beer, R.D., and Gallagher, J.C. (1999). Evolution and analysis of model CPGs for walking I. Dynamical modules. *J. Computational Neuroscience* **7**:99-118.

Churchland, P.S., Ramachandran, V.S. and Sejnowski, T.J. (1994). A critique of pure vision. In C. Koch and J.L. Davis (Eds.), *Large-Scale Neuronal Theories of the Brain* (pp. 23-60). MIT Press.

Clancey, W.J. (1997). *Situated Cognition*. Cambridge University Press.

Clark, A. (1997). *Being There: Putting Brain, Body and World Together Again*. MIT Press.

Clark, A. and Chalmers, David J. (1998). The extended mind. *Analysis* **58**:7-19.

Cliff, D. (1991). Computational neuroethology: A provisional manifesto. In J.A.Meyer and S. W. Wilson (Eds.), *From Animals to Animats: Proceedings of the First International Conference on Simulation of Adaptive Behavior* (pp. 29-39). MIT Press.

Cliff, D., Harvey, I. and Husbands, P. (1993). Explorations in evolutionary robotics. *Adaptive Behavior* **2**:73-110.

Di Paolo, E.A., (1998). An investigation into the evolution of communication. *Adaptive Behavior* **6**:285-324.

Di Paolo, E.A. (2000). Behavioral coordination, structural congruence and entrainment in a simulation of acoustically coupled agents. *Adaptive Behavior* **8**:25-46.

Di Paolo, E.A. and Harvey, I. (2003). Decisions and noise: The scope of evolutionary synthesis and dynamical analysis. *Adaptive Behavior* **11**:284-288.

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

Dreyfus, H.L. (1992). *What Computers Still Can't Do*. Original edition published in 1972. MIT Press.

Edelman, G.M., Reeke, G.N., Gall, W.E., Tononi, G., Williams, D and Sporns, O. (1992). Synthetic neural modeling applied to a real-world artifact. *Proc. Natl. Acad. Sci. USA* **89**:7267-7271.

Eliasmith, C. (1997). Computation and dynamical models of mind. *Minds and Machines* **7**:531-541.

Elman, J. L. (1995). Language as a dynamical system. In R. Port and T. van Gelder (Eds.), *Mind as Motion: Explorations in the Dynamics of Cognition* (pp. 195-225). MIT Press.

Floreano, D. and Mondada, F. (1996). Evolution of plastic neurocontrollers for situated agents. In P. Maes et al. (Eds.), *From Animals to Animats 4: Proceedings of the Fourth International Conference on Simulation of Adaptive Behavior* (pp. 402-410). MIT Press.

Gibson, J.J. (1979). *The Ecological Approach to Visual Perception*. Houghton Mifflin.

Gruau, F. (1995). Automatic definition of modular neural networks. *Adaptive Behavior* **2**:151-183.

Grossberg, S. (1969). Embedding fields: A theory of learning with physiological implications. *J. Mathematical Psychology* **6**:209-239.

Grush, R. (1997). Book review of Port and van Gelder's *Mind as Motion*. *Philosophical Psychology* 10:233-242.

Harnad, S. (1990). The symbol grounding problem. *Physica D* **42**:335-346.

Harvey, I. (1992/1996). Untimed and misrepresented: Connectionism and the computer metaphor. *AISB Quarterly* **96**:20-27. Originally appeared as University of Sussex Cognitive Science Research Paper 245 in 1992.

Harvey, I., Di Paolo, E., Wood, R. and Quinn, M. (2005). Evolutionary robotics: A new scientific tool for studying cognition. *Artificial Life* **11**:79-98.

Hayes, P.J., Ford, K.M and Agnew, N. (1994). On babies and bathwather: A cautionary tale. *AI Magazine* **15**(4):15-26.

Heideggar, M. (1962). *Being and Time*. Originally published in 1927. Harper and Row.

Husbands, P., Harvey, I. and Cliff, D. (1995). Circle in the round: State space attractors for evolved sighted robots. *Robotics and Autonomous Systems* **15**:83-106.

Hutchins, E. (1995). *Cognition in the Wild*. MIT Press.

Izhikevich, E.M. (2007). *Dynamical Systems in Neuroscience*. MIT Press.

Izquierdo-Torres, E. and Harvey, I. (2006). Learning on a continuum in evolved dynamical node networks. In L. Rocha et al. (Eds.), *Proceedings of the Tenth International Conference on the Simulation and Synthesis of Living Systems* (pp. 507-512). MIT Press.

Kelso, J.A.S. (1995). *Dynamic Patterns: The Self-Organization of Brain and Behavior*. MIT Press.

Kimura, M. and Nakano, R. (1998). Learning dynamical systems by recurrent neural networks from orbits. *Neural Networks* **11**:1589-1599.

Kirsch, D. (1991). Today the earwig, tomorrow man? *Artificial Intelligence* **47**:161-184.

Kodjabachian, J. and Meyer, J.-A. (1998). Evolution and development of neural controllers for locomotion, gradient-following, and obstacle avoidance in artificial insects. *IEEE Trans. Neural Networks* **9**:796-812.

Kugler, P.N, Kelso, J.A.S. and Turvey, M.T. (1980). On the concept of coordinative structures as dissipative structures. In G.E. Stelmach and J. Requin (Eds.), *Tutorials In Motor Behavior* (pp. 3-47). North Holland.

Kuznetsov, Y.A. (2004). *Elements of Applied Bifurcation Theory* (3rd Edition). Springer.

Lakoff, G. and Johnson, M. (1999). *Philosophy in the Flesh*. Basic Books.

Langton, C.G., Ed. (1989). *Artificial Life*. Addison-Wesley.

Maes, P., Ed. (1990). *Designing Autonomous Agents*. MIT Press.

Marocco D., Cangelosi A., Nolfi S. (2003). The emergence of communication in evolutionary robots. *Philosophical Transactions of the Royal Society London - A* **361**: 2397-2421.

Maturana, H.R. and Varela, F.J. (1980). *Autopoieis and Cognition: The Realization of the Living*. D. Reidel Publishing.

Merleau-Ponty, M. (1962). *Phenomenology of Perception*. Humanities Press.

Miller, G.F. and Todd, P.M. (1991). Exploring adaptive agency: I. Theory and methods for simulating the evolution of learning. In D.S. Touretzkey et al. (Eds.), *Connectionist Models: Proceedings of the 1990 Summer School* (pp. 65-80). Morgan Kaufman.

Moreno, A., Umerez, J. and Ibañez, J. (1997). Cognition and life: The autonomy of cognition. *Brain and Cognition* **34**:107-129.

Noë, A. (2004). *Action in Perception*. MIT Press.

Nolfi S. (2005). Emergence of communication in embodied agents: Co-adapting communicative and non-communicative behaviors. *Connection Science* **17**:231-24.

Nolfi, S. and Floreano, D. (2000). *Evolutionary Robotics*. MIT Press.

Pfeiffer, R. and Scheier, C. (1999). *Understanding Intelligence*. MIT Press.

Port, R.F. and van Gelder, T., Eds. (1995). *Mind as Motion: Explorations in the Dynamics of Cognition*. MIT Press.

Rosenschein, S.J. and Kaelbling, L.P. (1986). The synthesis of digital machines with provable epistemic properties. In J.Y. Halpern (Ed.), *Proceedings of the Conference on Theoretical Aspects of Reasoning about Knowledge* (pp. 83-98). Morgan Kaufmann.

Schöner, G., Dose, M. and Engels, C. (1995). Dynamics of behavior: Theory and applications for autonomous robot architectures. *Robotics and Autonomous Systems* **16**:213-245.

Schöner, G. and Kelso, J.A.S. (1988). Dynamic pattern generation in behavioral and neural systems. *Science* **239**:1513-1520.

Skarda, C.A. and Freeman, W.J. (1987). How brains make chaos in order to make sense of the world. *Behav. Brain Sci.* **10**:161-195.

Slocum, A.C., Downey, D.C. and Beer, R.D. (2000). Further experiments in the evolution of minimally-cognitive behavior: From perceiving affordances to selective attention. In J.-A. Meyer et al. (Eds.), *From Animals to Animats 5: Proceedings of the Sixth International Conference on Simulation of Adaptive Behavior* (pp. 430-439). MIT Press.

Smithers, T. (1995). Are autonomous agents information processing systems? In L. Steels and R. Brooks (Eds.), *The Artificial Life Route to Artificial Intelligence* (pp. 123-162). Lawrence Erlbaum.

Steels, L. (2003) Evolving grounded communication for robots. *Trends Cog. Sci.* **7**:308-312.

Stewart, J. (1996). Cognition = Life: Implications for higher-level cognition. *Behavioural Processes* **35**:311-326.

Strogatz, S.H. (1994). *Nonlinear dynamics and chaos*. Addison-Wesley.

Suchman, L.A. (1987). *Plans and Situated Actions*. Cambridge University Press.

Thelen, E. and Smith, L.B. (1994). *A Dynamic Systems Approach to the Development of Cognition and Action*. MIT Press.

Thelen, E., Schöner, G., Scheier, C. and Smith, L.B. (2001). The dynamics of embodiment: A field theory of infant preservative reaching. *Behav. Brain Sci.* **24**:1-86.

Tuci, E., Quinn, M. and Harvey, I. (2002). An evolutionary approach to the study of learning behavior using a robot-based model. *Adaptive Behavior* **10**:201-221.

Turner, J. Scott. (2000). *The Extended Organism: The Physiology of Animal-Built Structures*. Harvard University Press.

Turvey, M.T. (1990). Coordination. *American Psychologist* **45**:938-953.

Van Gelder, T. (1995). What might cognition be if not computation? *J. Philosophy* **91**:345-381.

Van Gelder, T. (1998). The dynamical hypothesis is cognitive science. *Behva. Brain Sci.* **21**:615-628.

Van Leeuwen, M. (2005). Questions for the dynamicist: The use of dynamical systems theory in the philosophy of cognition. *Minds and Machines* **15**:271-333.

Varela, F.J., Thompson, E. and Rosch, E. (1991). *The Embodied Mind*. MIT Press.

Vaughan, E., Di Paolo, E.A. and Harvey, I. (2004). The evolution of control and adaptation in a 3D powered passive dynamic walker. In J. Pollack et al. (Eds.), *Proceedings of the Ninth International Conference on the Simulation and Synthesis of Living Systems* (pp. 139-145). MIT Press.

Vera, A.H. and Simon, H.A. (1993). Situated action: A symbolic interpretation. *Cognitive Science* (with commentary and responses) **17**:7-48

Vickerstaff, R.J. and Di Paolo, E.A. (2005). Evolving neural models of path integration. *J. Exp. Biol.* **208**:3349-3366.

Walter, W.G. (1953). *The Living Brain*. W.W. Norton.

Ward, R. and Ward, R. (2006). Cognitive conflict without explicit conflict monitoring in a dynamical agent. *Neural Networks* **19**:1430-1436.

Warren, W.H. (2006). The dynamics of perception and action. *Psychological Review* **113**:358-389.

Webb, B. (2001). Can robots make good models of biological behavior? *Behav. Brain Sci.* **24**:1033-1050.

Werner, X and Dyer, M. (1992). Evolution of communication in artificial organisms. In C. Langton et al. (Eds.), *Artificial Life II* (pp. 659-687). Addison-Wesley.

Wheeler, M. (2005). *Reconstructing the Cognitive World*. The MIT Press.

Winograd, T. and Flores, F. (1986). *Understanding Computers and Cognition*. Ablex Publishing.

Yamuchi, B. and Beer, R.D. (1994). Sequential behavior and learning in evolved dynamical neural networks. *Adaptive Behavior* **2**:219-246.