COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science (2010) 1–36
Copyright © 2010 Cognitive Science Society, Inc. All rights reserved.
ISSN: 0364-0213 print / 1551-6709 online
DOI: 10.1111/j.1551-6709.2010.01151.x

Extending Dynamical Systems Theory to Model Embodied Cognition

Scott Hotton,[a] Jeff Yoshimi[b]

[a]*Department of Organismic and Evolutionary Biology, Harvard University*
[b]*School of Social Sciences, Humanities and Arts, University of California, Merced*

Received 15 September 2009; received in revised form 19 June 2010; accepted 23 June 2010

Abstract

We define a mathematical formalism based on the concept of an ''open dynamical system'' and show how it can be used to model embodied cognition. This formalism extends classical dynamical systems theory by distinguishing a ''total system'' (which models an agent in an environment) and an ''agent system'' (which models an agent by itself), and it includes tools for analyzing the collections of overlapping paths that occur in an embedded agent's state space. To illustrate the way this formalism can be applied, several neural network models are embedded in a simple model environment. Such phenomena as masking, perceptual ambiguity, and priming are then observed. We also use this formalism to reinterpret examples from the embodiment literature, arguing that it provides for a more thorough analysis of the relevant phenomena.

Keywords: Dynamical systems; Embodied cognition; Neural networks; Representation

1. Introduction

The embodied cognition tradition is by now a complex and far-reaching enterprise, spanning psychology, cognitive science, robotics, and philosophy, among other disciplines (see Chemero, 2009; Clark, 2008; Gibbs, 2006; and Noë, 2004).[1] The core idea is that cognition extends beyond the traditional boundaries of skin and skull, encompassing artifacts and features in the environment. Some have suggested doing away with the traditional concept of an internal representation altogether, focusing instead on an integrated animal environment system (the idea goes back at least to Gibson, 1986).

Proponents of the concept of an internal representation have mounted a wide-ranging counterattack (for review and an externalist response, see part 2 of Clark, 2008). Among

Correspondence should be sent to Jeff Yoshimi, School of Social Sciences, Humanities and Arts, 5200 North Lake Road, University of California, Merced, CA 95343. E-mail: jyoshimi@ucmerced.edu

other things, it has been pointed out: (1) that the brain processes information in a distinctive way, as compared with external artifacts like pen and paper or cell phones (see Adams & Aizawa, 2001, and Rupert, 2004), (2) that internal representations—which are enduring, mediating, and often amodal—maintain an important explanatory role in cognitive science (see Markman & Dietrich, 2000a), and (3) that the brain performs important control theoretic functions, often independently of environmental coupling (see Grush, 2003). Moreover, (4) there is a long-standing tradition in cognitive science and philosophy that assumes that the immediate causal basis of conscious experience is in the brain (in fact, Clark, a prominent exponent of the externalist position, has recently defended this idea against those who argue for an ''extended conscious mind;'' see Clark, 2009).

We describe an intermediate position in the embodied cognition debate, using a class of mathematical objects that we have called ''open dynamical systems'' (see Hotton & Yoshimi, 2010). An open dynamical system has a dynamical system to model an agent and a dynamical system to model an environment with an embodied agent. This makes it possible to compare a system's behavior when it is isolated from any environment with its behavior when it is embedded in a particular environment. In this way, internalist and externalist considerations can be accommodated in a unified framework. By taking this mixed approach, new forms of analysis are facilitated. In particular, we introduce tools for analyzing the complex, overlapping collections of paths that occur in the state space for embodied agents. The geometric and topological structure of these paths sheds light on the way embodied agents represent their environments over time.[2,3]

In this article, we generalize our previous work (Hotton & Yoshimi, 2010) on the subject, where our focus was on technical matters faced by the embodied research community. The main issue we addressed there was the fact that embodied cognition studies ''open'' interactions between agents and their environments, but often rely on dynamical systems, as traditionally defined, which are in a certain sense closed. This has led some theorists to use key concepts in an inconsistent way. For example, fixed points are allowed to ''move,'' and orbits are allowed to pass through each other, although that is not strictly speaking possible in a classical dynamical system. By formally defining open dynamical systems, we do not reject these ways of thinking, but rather provide them with a rigorous underpinning.

Also, in our previous work, we focused on a somewhat narrow class of continuous dynamical systems. We here generalize to a much broader class of dynamical systems.

In Sections 2 and 3, we define dynamical systems, open dynamical systems, and associated concepts. We also introduce tools for analyzing and classifying the paths that occur in an open dynamical system. The concept of a mental representation is discussed in Section 4. In Section 5, we use these tools to model embodied cognition. We embed a sensory network and a continuous Hopfield network, respectively, in a simple model environment, and show how, even in these simple examples, psychologically plausible forms of representational processes such as masking, perceptual ambiguity, and priming can be observed. In Section 6, we reinterpret a series of examples from the embodied cognition literature, arguing in each case that the relevant phenomena can be more completely analyzed using open dynamical systems.

2. Dynamical systems

Open dynamical systems are defined in terms of dynamical systems, so we begin with a brief overview of dynamical systems theory.[4] A dynamical system is a mathematical description of how things change with time. At any moment in time, a dynamical system is said to occupy a particular state. The set S of all possible states of a dynamical system is called its *state space*. Dynamical systems also have a *time space* T, which is the set of all moments in time.[5] One of the key properties of dynamical systems is that they are deterministic in the sense that the present state uniquely determines the states at all future times (though dynamical systems can exhibit complex behavior which, in practice, can be difficult to predict, e.g., via chaotic dynamics).

Abstractly, a dynamical system is a function of the form $\phi{:}S \times T \rightarrow S$. This function takes a state $s_0 \in S$ (which we think of as an initial condition) and a time $t \in T$ and returns the state the system will be in at time t starting from state s_0. This state can be written as $\phi(s_0,t)$. To be a dynamical system, the function ϕ must satisfy the two properties:

- There is a time $t_0 \in T$ such that for all states $s_0 \in S \quad \phi(s_0,t_0) = s_0$.
- For all states $s_0 \in S$ and all times $t_1,t_2 \in T \quad \phi(s_0,t_1 + t_2) = \phi(\phi(s_0,t_1),t_2)$.

The first property says that there is some time, t_0, that we take to be the present moment and that each state must be mapped to itself in the present moment. The second property says that future states are uniquely determined by the present state.[6]

Many mathematical models in science satisfy this abstract definition. For example, in many cases, the solution of a differential equation is a dynamical system,[7] as are many computer programs (where the state space is the set of discrete states of the machine running the program). Many mathematical models in cognitive science are dynamical systems in this sense, for example, many connectionist networks and artificial intelligence simulations.[8]

One advantage of the dynamical systems perspective is that it provides a set of tools that can be used to analyze a system, even when the solution to a differential equation cannot be expressed in terms of elementary functions. The basic idea is to focus on (and where possible, visualize) the set of possible behaviors of a system, and to classify those behaviors. To give a sense of how this works, we begin by (informally) defining relevant concepts. We then illustrate how these concepts can be used to analyze several classical dynamical systems. In the next section, we define open dynamical systems and show how the same tools and techniques that are used in analyzing classical dynamical systems can be adapted to the analysis of open dynamical systems.

The set of all points that can be reached from a given initial state is an *orbit*.[9] (In some contexts orbits are also called ''*trajectories*''). Intuitively, an orbit is one possible time evolution of a system, one possible behavior of the system over time. A collection of orbits for a dynamical system is a *phase portrait*; it shows possible evolutions, possible behaviors, for that system. (Of course, graphical depictions of phase portraits can only contain illustrative subsets of a system's orbits.) Phase portraits can convey an intuitive understanding of a dynamical system's structure (see Fig. 1 for three examples). A *parameter* is a quantity used

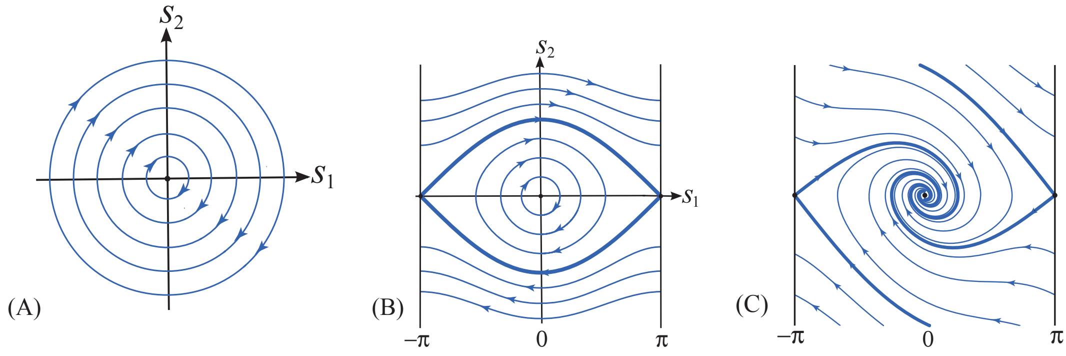

Fig. 1. (A) Phase portrait for a harmonic oscillator. The state space is a plane. The origin is a fixed point and the other orbits are circles. (B) Phase portrait for an undamped pendulum. The state space is a cylinder, which is displayed as a flat vertical strip whose edges are to be attached since $-\pi$ and π stand for the same angular displacement. The points (0,0) and ($\pm\pi$,0) are the fixed points of the dynamical system. Almost all of its orbits are simple closed curves, which either go around (0,0) or wrap around the cylinder. (C) Phase portrait for a damped pendulum. The point (0,0) is an attracting fixed point. ($\pm\pi$,0) is an unstable fixed point. Two orbits converge to ($\pm\pi$,0) and all of the other orbits converge to (0,0).

to specify how a dynamical system acts. A parameter's value is fixed as a dynamical system operates, but we can abstractly consider how changing the parameter's value alters the behavior of the dynamical system. For each parameter value, the phase portrait can have a distinctive form, so that we can speak of a ''parameterized family of dynamical systems.'' A *bifurcation* is a qualitative change that can occur when a parameter is varied, so that the phase portrait for the dynamical system before the bifurcation is qualitatively different from the phase portrait for the dynamical system after the bifurcation.

It is often possible to understand all of the behaviors of a dynamical system by classifying the orbits and analyzing them topologically and geometrically. We can begin by identifying the *exceptional* orbits in a phase portrait, which roughly means orbits that are different in form from their neighbors. Exceptional orbits can separate a phase portrait into collections of similar orbits. We refer to the collections of similar orbits that lie outside of exceptional orbits as ''isotypes,'' and we refer to the orbits within an isotype as ''generic orbits.''[10] The isotypes in a phase portrait illustrate the main classes of typical behavior for a dynamical system. Exceptional orbits, generic orbits, and isotypes are formally defined below in a way that applies to both open dynamical systems and classical dynamical systems.

Having distinguished the exceptional orbits and isotypes of generic orbits in a phase portrait, we can further refine the classification. One important type of orbit is a *fixed point* (a point which the dynamical system maps to itself, that is, a state p such that $\phi(p,t) = p$ for all t). An *attracting* fixed point is one where all orbits sufficiently close to it converge to it. An *unstable* fixed point is one where some nearby orbits move further away from it without returning. A fixed point need not be attracting or unstable. An attracting fixed point is the simplest type of an attractor. More generally an *attractor* is a set where all nearby orbits move toward it. Attractors correspond to the behaviors that a dynamical system tends to exhibit over time. The orbits within an attractor are often exceptional, but the attractor still informs us about the typical behavior of the system, because neighboring orbits, which tend to be generic, stay near the attractor once they are sufficiently close to it (even though they may not actually reach it in a finite amount of time). The region of state space containing all of the states that tend toward an attractor make up its *basin of attraction*. Orbits in a basin of attraction are *basin orbits*. Some attractors are *chaotic* and have sensitive dependence on initial conditions, so in practice their precise long-term behavior is unpredictable.

We conclude this section with three classic examples that illustrate how these concepts can be used to analyze dynamical systems.

The first example is a harmonic oscillator. A spring that obeys Hooke's law is called a harmonic oscillator because it oscillates with a single frequency, a fact which is useful in making musical instruments. By Hooke's law, the force pulling a stretched spring back to its equilibrium length produces an acceleration that is proportional to how far the spring is stretched from its equilibrium length. If we let s_1 stand for the spring's displacement as a function of time, then, in appropriate units of measure, Hooke's law is expressed by the differential equation $\ddot{s_1} = -s_1$.

To demonstrate how solutions to differential equations can satisfy the abstract definition of a dynamical system, we give the solution to $\ddot{s_1} = -s_1$ here. As this is a second-order differential equation, to solve it we need to know both how far the spring is stretched and how fast it is set in motion at the initial moment, which we take to be $t_0 = 0$. We let s_2 stand for the spring's velocity as a function of time, $x_0 = s_1(0)$ stand for the initial displacement, and $v_0 = s_2(0)$ stand for the initial velocity. The state of a harmonic oscillator at any moment in time is given by the pair $(s_1(t),s_2(t))$, so the state space is the plane, $\mathbf{R}^2$. The solution can be expressed as:

$$\begin{pmatrix} s_1(t) \\ s_2(t) \end{pmatrix} = \begin{pmatrix} x_0\cos(t) + v_0\sin(t) \\ -x_0\sin(t) + v_0\cos(t) \end{pmatrix} = \begin{pmatrix} \cos(t) & \sin(t) \\ -\sin(t) & \cos(t) \end{pmatrix} \begin{pmatrix} x_0 \\ v_0 \end{pmatrix} = \phi(x_0, v_0, t)$$

It can be checked that $\dot{s_1}(t) = s_2(t)$ and $\ddot{s_1}(t) = -s_1(t)$, so this expression does indeed solve the differential equation.

The equation also shows that the solution can be written as the product of a matrix times a vector. The matrix represents a rotation by $-t$ radians about the origin of $\mathbf{R}^2$ and the vector is the initial condition (x_0,v_0). The solution at any given moment in time is a function that depends on the values of the initial conditions (x_0,v_0) and on the time t. We denote this function by ϕ. We can see how ϕ satisfies the two properties for a function to be a dynamical system. For the first property, recall that we took $t_0 = 0$ to be the initial moment and indeed $\phi(x_0,v_0,0) = (x_0,v_0)$ because a rotation by 0 radians fixes the plane. The fact that ϕ satisfies the second property is a consequence of the composition rule for rotations about a common center. Rotating about the origin by $-(t_1 + t_2)$ radians is the same as rotating about the origin by $-t_1$ radians and then by $-t_2$ radians.

A phase portrait for the harmonic oscillator is shown in Fig. 1A. We can see intuitively in the figure how the spring behaves. It is either at rest or oscillating at a particular frequency. These two characteristic behaviors correspond to:

- An exceptional orbit. This is the origin, which corresponds to a spring at rest in its equilibrium length. The exceptional orbit is a fixed point, but it is neither attracting nor unstable (this type of fixed point is called a ‘‘center’’).
- An isotype of generic orbits. These are the concentric circles, which correspond to oscillations with a fixed nonzero frequency. The radius of the circles is the amplitude of the corresponding oscillation.

The generic orbits characterize the behavior we expect for an ideal harmonic oscillator because the slightest amount of energy will set it oscillating.

The second example is a frictionless pendulum, which is an idealized model of a pendulum that, once it is set in motion, continues to move forever. The bob of a pendulum is constrained to move along a circle that is perpendicular to the ground and whose center is the pendulum’s pivot. We let s_1 stand for the angular displacement of the bob from the straight down direction and s_2 stand for the angular velocity. The set of all possible values for s_1 is a circle, that is, the interval $[-\pi,\pi]$ with the two end points identified with each other. The set of all possible values of s_2 forms a line, which corresponds to the fact that the pendulum can rotate clockwise or counterclockwise with arbitrarily large or small angular velocities. The state space for the pendulum is thus the Cartesian product of a circle with a line, which forms a cylinder. Each point on the cylinder corresponds to a unique angular position and angular velocity for the pendulum.

A phase portrait for the frictionless pendulum is shown in Fig. 1B. We can again see intuitively in the figure how the system behaves. In this case, there are four exceptional orbits and three isotypes.

- The two fixed points are exceptional. The fixed point at the origin corresponds to the pendulum at rest. The unstable fixed point at $(\pm\pi,0)$ corresponds to the pendulum being balanced in the upright position.
- The other two exceptional orbits are curves connecting the unstable fixed point to itself (‘‘homoclinic connections’’). These orbits correspond to the pendulum falling from a nearly upright position and swinging back up to a nearly upright position in either the clockwise or counterclockwise direction.
- The three isotypes: a central region of simple closed curves centered about the origin, which corresponds to the pendulum swinging back and forth forever, and the upper and lower regions of simple closed curves going around the cylinder, which correspond to the pendulum rotating clockwise or counterclockwise forever.

The third example is a damped pendulum, a variant on the last example, which has been made more realistic by the addition of a frictional force. The state space is still a cylinder, but the phase portrait has changed to reflect the differences in its behavior introduced by friction. A phase portrait for a damped pendulum is shown in Fig. 1C. In this case there are six exceptional orbits and two isotypes.

- The fixed points are the same as with the frictionless pendulum (the origin and $(\pm\pi,0)$). The origin is now attracting. Instead of being surrounded by concentric simple closed curves, the orbits near (0,0) tend to spiral toward it. This corresponds to the fact that a pendulum subject to frictional forces will lose energy and swing with an amplitude that decreases with time.

- Two of the exceptional orbits correspond to the pendulum rotating clockwise or counterclockwise and slowing down to a nearly upright position.
- The two remaining exceptional orbits correspond to the pendulum falling either clockwise or counterclockwise from a nearly upright position and swinging with a decaying amplitude to rest. These two exceptional orbits along with the two fixed points separate the two isotypes.
- The two isotypes correspond to the pendulum rotating either clockwise or counterclockwise and then swinging to rest.

These two isotypes correspond to the two types of behaviors that will be observed in practice for a pendulum. If the damped pendulum's initial angular velocity is small enough, a time span of rotation will not be observed and it would be difficult for a casual observer to tell which of the two isotypes the pendulum was following.

In the transition from the damped to the frictionless pendulum, we see an example of a bifurcation. The level of damping can be thought of as a parameter which, when varied, changes the overall behavior of the pendulum. For most values of this parameter, the phase portrait remains qualitatively the same, in that there are the same number of exceptional orbits and isotypes and they contain orbits with the same topology. What mainly changes is the rate at which the pendulum approaches its rest state. However, when damping is reduced to zero, a qualitative change occurs; there are no longer six exceptional orbits and two isotypes, but four exceptional orbits and three isotypes.

3. Open dynamical systems

We now add more structure to the concept of a dynamical system to model interactions between an agent and environment. In doing so, we will see that the techniques used above to characterize the behaviors of a dynamical system can be carried over to the case of an open dynamical system.

The basic concept of an open dynamical system is simple: It is a compound system consisting of (1) a dynamical system modeling an agent by itself, (2) a dynamical system modeling an agent in an environment, and (3) some machinery for showing how they are related. For a visual overview of the structures defined here, see Fig. 2. We begin by considering the structures on the right side of the figure.

A *total space* S_τ is the set of possible states of an environment with at least one agent embedded in it. An *agent space* S_α is the set of possible states for a particular agent. The environment and the agent influence each other's behavior, and we assume that their behavior together can be specified with a single dynamical system, a *total system* ϕ_τ on S_τ, somewhat like Gibson's animal-environment system. We assume that for each state of the total system there is a corresponding state for the agent. This can be captured by a map,

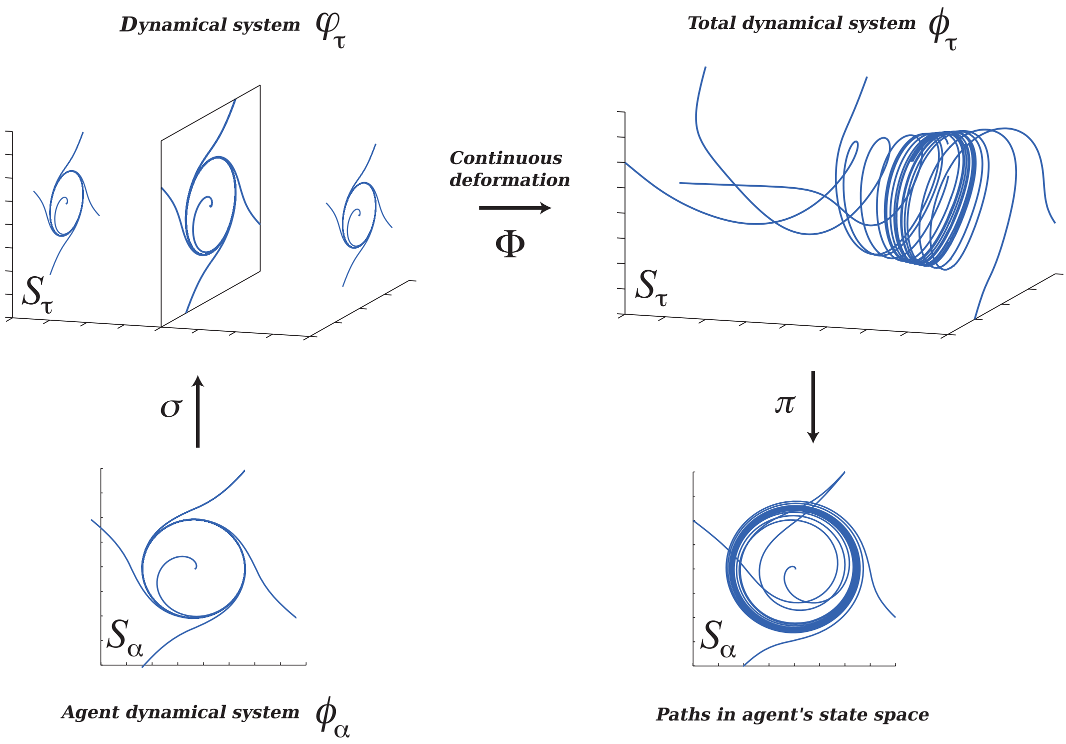
Dynamical system φ_τ
Total dynamical system ϕ_τ
S_τ
S_τ
Continuous deformation
Φ
σ
π
S_α
S_α
Agent dynamical system ϕ_α
Paths in agent's state space

Fig. 2. Schematic of an open dynamical system. Lower left: an agent dynamical system ϕ_α defined on the agent state space S_α, which shows how the agent behaves in isolation. Upper left: a system φ_τ defined on the total space S_τ, which shows how the decoupled agent system and the total system are related. Upper right: a total dynamical system ϕ_τ defined on the total space S_τ obtained by a continuous deformation of φ_τ, which describes the behavior of the agent in its environment. Lower right: projection of the orbits of the total system to the agent space (i.e., the paths), which corresponds to the behavior of an agent in a particular environment (an open system with open dynamics). Note the contrast between the phase portrait for the isolated agent (lower left) and the open phase portrait for the embedded agent (lower right). The paths in the open phase portrait overlap, which makes them harder to analyze.

$\pi{:}S_\tau \rightarrow S_\alpha$, from the total state space to the agent's state space. The map π is not in general one to one. For example, many changes can occur outside of an agent's field of view that will not immediately affect the agent's state. We also assume there is a map, $\sigma{:}S_\alpha \rightarrow S_\tau$, which places the agent space in the total space.[11]

At this point, it is important to flag a potential confusion. The map π is a function from the total space to the agent space, which could suggest that we are only interested in couplings from an environment to an agent, as opposed to two way couplings where agents and environments both affect one another. However, this is not the way π should be understood; π is not a coupling term. The total system ϕ_τ describes the dynamics of an agent and an environment (and perhaps multiple other agents) combined together, all of which can be tightly coupled to one another. The projection map π has nothing to do with such couplings; it is simply a tool for ''viewing'' what happens inside an agent. Although the total systems considered below only involve one direction of coupling, that is a just matter of convenience. Nothing in this framework precludes an analysis of two-way couplings, and we plan to analyze such systems in future studies.

We want to be able to compare the behavior of an agent when it is in an environment with the behavior of an agent when it is isolated from any environment (e.g., the two bottom panels of Fig. 2). To do this, we assume the agent has its own dynamical system, an *agent system* ϕ_α, which corresponds to the ''intrinsic'' or ''closed'' dynamics of the agent—its behavior when it is not in an environment (bottom left of Fig. 2). To compare an agent's closed and open dynamics, we begin by (roughly speaking) ''placing'' the phase portrait corresponding to the closed dynamics ϕ_α within a phase portrait φ_τ on the total space (see the left side of Fig. 2). More precisely, φ_τ is a dynamical system on the total space such that $\sigma(S_\alpha)$ is invariant under φ_τ and such that the restriction of φ_τ to $\sigma(S_\alpha)$ is a dynamical system equivalent to ϕ_α.[12] We can think of φ_τ as a total system corresponding to the decoupled agent.

To understand the relationship between the isolated agent system and the embedded agent system, we assume that there is a way to continuously deform the total system, ϕ_τ, into the total system corresponding to the decoupled agent, φ_τ (the two top panels of Fig. 2). This is done using a ''homotopy,'' a function that describes continuous deformations. In this case, the homotopy is a function $\Phi: S_\tau \times T \times [0,1] \rightarrow S_\tau$. The last argument (a value between 0 and 1) describes the various stages of the deformation: 1 corresponds to the total system; 0 corresponds to the decoupled total system; intermediate values correspond to intermediate stages of the deformation. That is, for any $s \in S_\tau$ and $t \in T$ we have

$$\Phi(s,t,0) = \varphi_\tau(s,t)$$
$$\Phi(s,t,1) = \phi_\tau(s,t)$$

Altogether, an open dynamical system as we have defined it consists of three spaces, S_α, S_τ, T, and four maps, π, σ, ϕ_α, Φ (the dynamical systems ϕ_τ and φ_τ are restrictions of Φ). If these seven objects are consistent with the assumptions described above, then we have an *open dynamical system*.

We now define open analogs of various structures that occur in classical dynamical systems, so that they can be put to use in analyzing open dynamical systems. The analog of an orbit in an open dynamical system is a *path*, which is the image of an orbit of ϕ_τ under the projection π (e.g., the projection of the behavior of an animal-world system down to the state space of the animal). That is, a path is a set of states along which the agent's state travels when embedded in an environment.

As paths can cross themselves, they can be fairly exotic when compared with what occurs in the phase portrait of a classical dynamical system. In particular, they possess topologies such as figure-eights, bouquets of arcs (collections of arcs joined at a single point), and Lissajous figures, which are impossible for orbits in a dynamical system.

The collection of all the paths in an agent space is the *open dynamics* of the agent and we will refer to agents with open dynamics as *open systems*. A collection of paths in an agent space is an *open phase portrait*. An open phase portrait helps show what happens inside of the agent when it is embedded in an environment (see the lower right of Fig. 2). Note that a formal distinction is made between ''open dynamical systems'' and ''open systems.'' A full septuple as described above is an open dynamical system. An agent whose future behavior is given by an open dynamical system is an open system.

Open systems are well suited to analyzing embodied cognition. Just as we can analyze the behavior of a classical dynamical system by considering different types of orbits in its phase portrait, we can analyze the behavior of an open system by considering different types of paths in its open phase portrait. We analyze representational processes in open systems in Section 5.

There is an important difference between a classical and an open phase portrait. The *orbits* of a classical phase portrait cannot cross (if they did cross, multiple futures would be possible from the cross-point, which would violate the assumption that initial conditions have unique futures). In an open phase portrait, by contrast, *paths* can cross (from such a cross point multiple futures are possible, depending on what happens in the environment). An example is shown in the lower right panel of Fig. 2. Whereas classical phase portraits, for example, those shown in Fig. 1, contain neatly separated collections of orbits, open phase portraits can contain dense tangles of overlapping projected orbits. This makes open phase portraits more complex and harder to analyze than classical phase portraits. But we will see that it is still possible to apply the techniques described above in such cases.

Three additional comments on these definitions are as follows. First, the total space S_τ will often be the Cartesian product of an *environment state space* S_ϵ and an agent state space (so that $S_\tau = S_\epsilon \times S_\alpha$). However, we do not require the total state space to have this form. For modeling embodied cognition, we only need to assume that for every state of the total system there is a unique agent state.[13]

Second, because of the extra assumptions that go into the definition of an open dynamical system, it might seem that the category of open dynamical systems is more restrictive than the category of dynamical systems, analogous to how, for example, magnolias form a more restrictive biological clade than flowering plants. However, strictly speaking, open dynamical systems are more general, because a classical dynamical system can be viewed as an open dynamical system if we let the total state space be the same as the agent's state space and let the projection and embedding be the identity map.

Third, note that this framework can be viewed as a generalization of two prominent accounts of embodiment in the literature (Beer, 2000, 2003; Warren, 2006) to a more abstract mathematical framework. Warren describes ''behavioral dynamics,'' whereby ''the agent and the environment can be treated as a pair of mutually coupled dynamical systems'' (see his Fig. 4, p. 367). Warren's account involves an agent dynamical system, an environmental dynamical system, and coupling terms between the two (''effector functions'' and ''information functions''). If we let Warren's agent systems be ''agent systems'' in our sense, and if we take the systems defined by Warren's coupled agent and environment systems to be ''total systems'' in our sense, then his account is an instance of ours. Beer does not develop an explicit framework, although what he has in mind is clearly quite similar to Warren's: ''an agent's nervous system, its body and its environment are each described as dynamical systems, which are coupled'' (p. 212). Thus, it is plausible that Beer's embodied agents can, like Warren's, be viewed as instances of open dynamical systems in our sense.

In the pendulum examples of the previous section, we saw that one virtue of dynamical systems theory is that even when a set of equations does not have a simple solution in terms of elementary functions, it is often still possible to classify the orbits in the system's phase portrait and thereby gain an understanding of its behavior. This was done by first identifying the ''exceptional orbits'' in a phase portrait, and then considering the collections or ''isotypes'' of ''generic orbits'' between these. We can now define these concepts from the standpoint of paths in an agent space. A generic path is, intuitively, a projection of an orbit of the total system that is part of a collection of orbits in the total system, all of which project to paths which do the same thing. Note that the preimage of a path, p, under the projection π must contain at least one orbit of ϕ_τ but it is possible that it could contain more than one orbit. Formally, a path p is *generic* if for some orbit, q, in its preimage there is a connected open set $U \subset S_\tau$, which contains q and orbits aside from q, such that every orbit contained in U projects to a path with the same topology as p. If a path is not generic, then it is an *exceptional path*. (Note that these definitions can also be applied to classical dynamical systems, because, as noted above, any classical dynamical system can be thought of as an open dynamical system where the total space is the agent space and the projection and embedding maps are the identity map.)

Isotypes are, intuitively, collections of generic paths that are all next to each other in the agent space, and similar in form. We can formally define isotypes as follows. Let $p_1,p_2 \subset S_\alpha$ be a pair of generic paths. If there is an isotopy between p_1 and p_2 (a continuous deformation which preserves the topology of the paths) such that every intermediate stage is also a generic path, then we say the paths p_1,p_2 are *isotypic*; a collection of isotypic paths is an *isotype*.

As we saw in the examples above (the oscillator; the damped and undamped pendulum), within an isotype, the behavior of a dynamical system is qualitatively the same when initial conditions change slightly. Being isotypic is an equivalence relation on the collection of generic paths in the agent space S_α, so that the isotypes of a system form a classification of the typical behaviors of that system (e.g., the typical forms of representational process for an embodied agent). The preimage under the projection π of the union of all of the paths of an isotype is an *isotype cover*. Isotype covers are useful in analyzing open dynamical systems because if one finds all the isotype covers in the total space, this implies that one has found all of the isotypes in the agent space.

Attractors—exceptional orbits that shed light on the observable behavior of a system since ''basin orbits'' approach them over time—also have analogs in open dynamical systems. The open dynamical system analog of attractors and basin orbits are simply the projections of attractor orbits and basin orbits from the total space down to the agent space. We call the projected attractor orbits ''*attracting paths*,'' and the projected basin orbits ''*trapped paths*'' (this captures the notion that they are ''trapped'' by the attracting paths). A formal definition of open dynamical systems analogs of attractors and basin orbits is given in the Appendix. When there are no attractors, then none of the paths are trapped. When there are attractors then, aside from some rather exceptional cases, all of the paths of an isotype will either be trapped or not trapped. When all of the paths of an isotype are trapped, we say that the ''*isotype is trapped*'' or that we have a ''*trapped isotype*.'' Trapped isotypes correspond to observable behaviors of open systems whose total systems have attractors.

4. Representation

Paths in an open phase portrait will be interpreted as representational processes below. This raises the question of what a ''representation'' is, such that we can describe paths as ''representational processes.'' The first thing to note is that this question is, strictly speaking, independent of the question of the utility of open dynamical systems, since this framework can be applied both in contexts where it is appropriate to refer to agent states as ''representations,'' and in contexts where this is not appropriate (i.e., in some cases there may be interesting agent dynamics that are not best regarded as ''representational dynamics'').

However, we do believe that in some cases, the concept of a representation has explanatory value (as we discuss in Section 6), and that in such cases, the paths of an open phase portrait can be interpreted as representational process. Moreover, we agree with Markman and Dietrich when they say that ''cognitive science needs multiple approaches to representation … No one representational format can handle all of these levels [of psychological data]'' (Markman & Dietrich, 2000b, p. 162). Insofar as there are distinct types of representation, there are correspondingly distinct kinds of representational dynamics that can be analyzed using this framework. In this paper, we focus on two.

First, we consider representations in the sense of indicators, internal states that carry information about the presence of some object in the environment. The basic idea is that a state s is a representation of object o if s occurs in the presence of o. For example, a red-light detector's on-state is a representation of red lights because it occurs when red-lights are present. This account runs in to numerous problems, and there is a whole industry devoted to properly defining the concept of a mental representation in this sense (notable philosophers working in this tradition are Fred Dretske and Ruth Millikan; for an overview, see Neander, 2004). In the specific examples considered in the next section, we assume a simple indicator view of representation, whereby the states the system occupies during prolonged exposure to external objects correspond to representations of those objects. In particular, we consider neural networks representing objects that pass the agent at slow and fast speeds, and mathematical models of biological cells that can represent the presence of other cells within a developing organism.

Second, we consider mental representations in a more complex sense (which builds on but is not equivalent to the indicator view), associated with the long-standing debate in cognitive science concerning what computational framework is most appropriate to cognitive research. The primary contenders until recently were symbolic AI and connectionism, each of which defends a particular view of mental representation. Common to these views is what Markman and Dietrich call a ''mediating state,'' where mediating states are ''internal states of a system that carry information which is used by the system in the furtherance of its goals'' (Markman & Dietrich, 2000b, p. 140). One value of this concept is that it is generic, ''It is intended to capture something that all cognitive scientists can agree to; namely, that there is internal information used by systems (organisms, for example) that mediates between environmental information coming in and behavior going out'' (p. 145). We consider dynamics of representations qua mediating states in Section 6 below, where we focus on insect navigation, agents playing the game *Scrabble*, humans finding there way around in complex environments, and navigators on a ship.

5. Examples

In this section, we describe several examples that show how open dynamical systems can be used to model representational processes in embodied agents. We define a simple environment and embed two simple neural network agents in this environment. The first network is extremely simple (a sensory network consisting only of two input nodes), which helps illustrate basic concepts. All of its isotypes can be found, so that we have an essentially complete understanding of the relationship between the agent and its environment. The second network is a continuous Hopfield network, which displays more complex behavior, because of its intrinsic dynamics. The interplay of intrinsic and environmental dynamics in the second example produce representational processes that are consistent with various observed perceptual phenomena, including masking, priming, and ambiguous perception, which suggests that open systems could be useful models in future research.[14] We also briefly consider a third example, from the field of biological development, whereby a cell in a developing organism represents (in the sense of indicating) neighboring cells.

We begin by defining an environmental dynamical system on an environment state space S_ϵ, which will be used for the two neural network examples. The model environment involves two objects that periodically come into and go out of view for an agent. We model this by placing the agent and two objects on a circle (see Fig. 3). The circle will be parame-

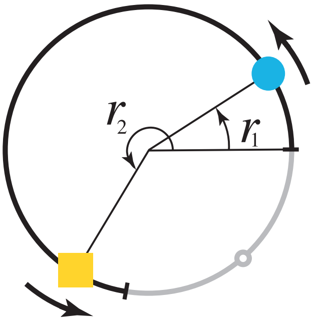

Fig. 3. A schematic of the environment. The agent is fixed on the circle at 300° (the gray dot) and its field of view (the gray arc) extends 60° in both directions. The objects travel around the circle. Object 1 is symbolized with a blue disk and object 2 is symbolized with a yellow square. The angular position of object 1 is given by r_1 and the angular position of object 2 is given by r_2.

terized by angle measured in degrees. The position of the agent is fixed at 300° and we assume that it has a ''field of view,'' which extends 60° in either direction. When an object enters this field of view, it stimulates the agent (that is, adds activation to one of its neurons, in a manner described below). We will limit ourselves to cases where the two objects travel along the circle at a fixed angular velocity.

The position of the *m*th object on the circle will be denoted by r_m. Thus, with two objects each moving at its own fixed angular velocity, the state of the environment over time is given by the ordered pair of angles (r_1,r_2). So the state space for the environment is the Cartesian product of a circle with itself. This forms a 2-torus, $\mathbf{T}^2$. In Figs. 4 and 6 below, we illustrate the dynamics on the 2-torus by using the Cartesian product of the interval [0°,360°] with itself. This forms a square. Since 0° and 360° stand for the same angular position, the opposite sides of the square need to be identified with each other to get the state space for the environment.

We denote the angular velocity of the *m*th object by v_m. We do not impose any limit on how fast the objects can go around the circle so that v_m can have any real value. The velocities of the objects together form a velocity vector (v_1,v_2). This gives us a dynamical system on the environment state space S_ϵ.

$$\phi_\epsilon(r_1, r_2, t) = (r_1, r_2) + t(v_1, v_2)$$

The behavior of this type of dynamical system is fairly simple. It is either periodic or quasi-periodic depending on the values of (v_1,v_2), which means, roughly, that its phase portrait will always consist of a family of parallel orbits traveling around the torus. We focus on the case where the two objects move at the same speed and maintain a fixed separation between one another as they travel around the circle. The phase portrait for this system consists of parallel orbits, which can be depicted as diagonal lines in the illustration of S_ϵ (see Fig. 4). Note that each orbit in the figure is actually a simple closed curve and that the environmental dynamical system is periodic over time.

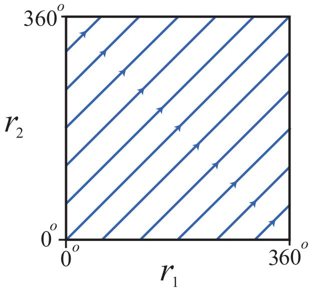

Fig. 4. Phase portrait for the environment dynamical system in which two objects move at the same speed along the circle. Since 0° and 360° stand for the same angle, the opposite sides of the square are to be identified with each other. The orbits that go into the top side come back from the bottom side. The orbits that go into the right side come back from left side. There are six orbits shown in the figure. Each orbit corresponds to a fixed angular separation between the objects.

We now use the techniques described above to analyze paths in the agent space of neural networks as representational processes. Following standard practice, we take the state space of a neural network to be its activation space (the set of possible patterns of activity across its nodes) and treat the weights between nodes as parameters that can vary during periods of learning but otherwise remain fixed when the network is in operation (see Churchland & Sejnowski, 1992). So the agent space S_α of a neural network is taken to be its activation space. The neural network's update rules define an agent dynamical system (an ''intrinsic'' or ''closed'' dynamics) $\phi_\alpha{:}S_\alpha \times T \;\rightarrow\; S_\alpha$.

We link the neural networks to the environment by having each object in the environment apply a scaled amount of activation to the corresponding node. More precisely, we let $d(r_m)$ be the angular distance between the mth object (positioned at r_m) and the agent (positioned at 300°) and we define a ''scaling function:''

$$f(d) = \begin{cases} 1 - d/60 & \text{if} \quad 0 \le d \le 60 \\ 0 & \text{otherwise} \end{cases}.$$

The value $f(d(r_m))$ is applied to the activation level, x_m, of the corresponding input node.

We can construct an open dynamical system by coupling a simple sensory network to the environment described above. This sensory network has no connections; it consists only of a pair of input nodes, which respond passively to the environment (see Fig. 5).

We can demonstrate that this system satisfies the formal definition of an open dynamical system, as follows. Let the nodes in this example take any value in [0,1]. Then the agent state space, S_α, is the square $[0,1]^2$ and agent dynamical system, ϕ_α, is just the identity map on $[0,1]^2$. In this case, we take the total state space, S_τ, to be the environment state space, S_ϵ, and the total dynamical system, ϕ_τ to be the environment dynamical system ϕ_ϵ. The maps π and σ will be

$$\pi(r_1, r_2) = (f(d(r_1)),\ f(d(r_2)))$$
$$\sigma(x_1, x_2) = (240,\ 240) + 60(x_1, x_2)$$

The dynamical system φ_τ is

$$\varphi_\tau(r_1, r_2, t) = (r_1, r_2)$$

Clearly $\sigma([0,1]^2) = [240°,300°]^2 \subset S_\tau$ is invariant under φ_τ and the restriction of φ_τ to $\sigma(S_\alpha)$ is the identity map. Therefore, on the invariant set $\sigma(S_\alpha)$ the dynamical system φ_τ is equivalent to ϕ_α. And finally, we specify the continuous map

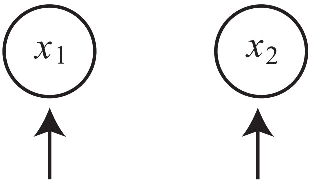

Fig. 5. Topology of the simple object detector. The two nodes are independent from each other and receive input from the environment.

$$\Phi(r_1, r_2, t, \mu) = (r_1, r_2) + \mu t(v_1, v_2)$$

It is also clear that $\Phi(r_1,r_2,t,0) = \varphi_\tau(r_1,r_2,t)$ and $\Phi(r_1,r_2,t,1) = \phi_\tau(r_1,r_2,t)$. In this way, we have an open dynamical system model for this simple neural network embedded in the circle world.

We interpret network states as representations by using a simple local coding scheme whereby the degree of activation of the first node represents (in the sense of indicating; see Section 4) the distance between the agent and the first object, and similarly with the second node and the second object. On this interpretation, points on the bottom left corner of S_α correspond to no objects being in view (which we can think of as a representation of no objects), points on the bottom and left edges of S_α correspond to representations of one object only (since one object is not in view), points on the right and top edges correspond to ''maximal'' representations of one object (one object is on top of the agent), and the top right corner corresponds to a maximal representation of both objects (both objects are on top of the agent). Points in the interior correspond to representations of both objects at various intermediate distances from the agent.

Paths can then be interpreted as representational processes. Paths ''bouncing'' off the right or upper edge of S_α correspond to an object passing by the agent, that is, the network's nodes reach their maximum value and decline. Paths arriving at or leaving the left or lower edge correspond to an object coming in to or going out of view, that is, declining in value to zero. These concepts are used to interpret the path diagrams below.

This open dynamical system has five isotypes, which we can think of as five types of representational process relative to five types of environmental configuration. It also has a single exceptional path (a representational processes that will rarely occur). In Fig. 6, we show the five isotype covers (the collections of orbits in the environment space that project to paths in an isotype). We also show three sample generic paths, and sample environmental configurations corresponding to each of these generic paths.

We can think of these different types of representational process by focusing on the topology of the paths, which display three distinct kinds of topology. (The exceptional path has the same topology as one of the isotypes, and there are two pairs of isotypes whose paths differ only in their orientation). These three kinds of topology for generic orbits, plus the exceptional orbit, correspond to four distinct ways the objects can pass the agent, and they result in four types of representational process.

1. The path for the isotype whose isotype cover is shown in green has the topology of an arc that has two straight segments (for example, the top path/environment configuration in Fig. 6). This is an example where the preimage of a path under π contains more than one orbit. On this path only a single object is represented at a time. As each object comes into view and passes by the agent, half of the arc is traced out as the corresponding node reaches its maximum value of 1 and goes back to 0.
2. The paths of the two isotypes whose isotype covers are shown in the two shades of blue have the topology of a loop with two arcs attached (for example, the middle path/environment configuration in Fig. 6). For these paths, the objects are closer together,

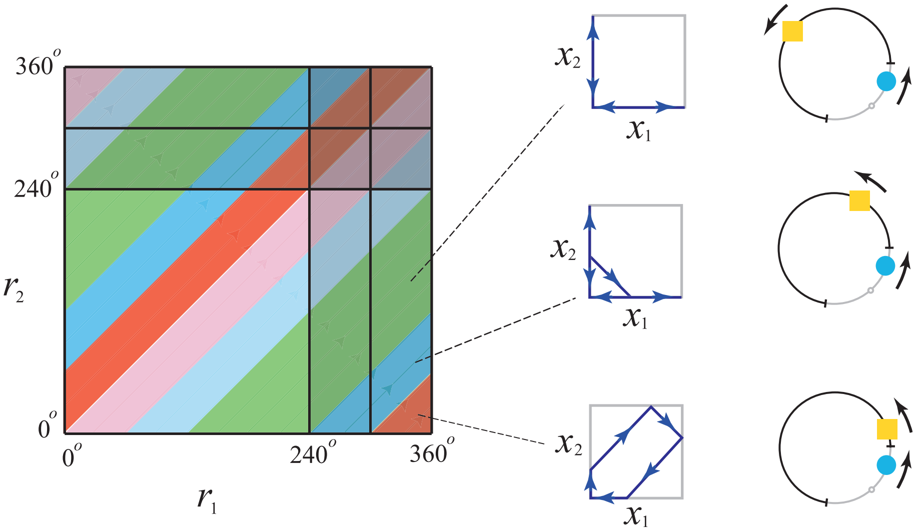

Fig. 6. The large square on the left depicts the state space $S_\tau = S_\epsilon$ for the circle world. The isotype covers of the agent system are shown in different colors. When the objects are in the dark shaded regions toward the top or right-hand sides of S_τ at least one of the agent's two sensors is activated. The pictures to the right show sample paths that occur in the agent state space S_α for three of the isotypes and configurations of objects in the circle world that could produce those paths.

so that the agent represents one object by itself, then the two objects together, then the other object by itself. First, one object comes into view and passes by the agent so that one node goes to its maximum value and begins to decrease. As that is happening, the other object comes in view so that for a time both nodes are active and the path enters the interior of S_α. The paths of these two isotypes are made up of the same points, but their orientation is reversed. For the light blue isotype cover, object 1 comes into view first. For the dark blue isotype cover, object 2 comes into view first.

3. The paths of the isotypes whose isotype covers are shown in shades of red have the topology of a circle (for example, the bottom path/environment configuration in Fig. 6). On these paths, the objects are even closer together so that soon after one object has come into view (and, unlike the case above, before that object passes by the agent) the other object is reaching the top or left edge. Again, there are two isotypes whose paths are made up of the same set of points but with orientation reversed.
4. The exceptional path (not shown) is a straight line segment, which connects (0,0) and (1,1). In this case, both objects come into view together, pass the agent together, and fade out of view together. That is, the agent represents no object, then both objects as they approach and pass the agent. During this time, both nodes go to their maximum value and decrease back to 0 simultaneously.

Our next neural network is a two-node continuous Hopfield network (this has been analyzed in greater detail in Hotton and Yoshimi, 2010). In this example, the neurons are coupled to each other (see Fig. 7) and they have their own dynamical properties. This makes the path shapes more complex and more interesting psychologically.[15] Because of the additional complexity associated with the intrinsic dynamics of the model neurons, chaotic behaviors can emerge in this example, and an enumeration of the isotypes is no longer feasible. This is analogous to the situation in dynamical systems theory where it is often possible to classify completely the orbits of a continuous dynamical system on a two-dimensional manifold, whereas it can be practically impossible to do the same for continuous dynamical systems in three or more dimensions.

The state space for a Hopfield network with N neurons is $\mathbf{R}^N$ and we denote the state of the neurons by $(x_1,\ldots,x_N)$. The intrinsic dynamics of a continuous Hopfield network is given by the system of differential equations

$$\dot{x_j} = -x_j/R_j + \sum_{i=1,\neq j}^{N} w_{ji}\, g(x_i) \quad j = 1,\ldots,N$$

where $g(x_i) = (2/\pi)\arctan(\pi\lambda x_i/2)$ is shaped as a sigmoidal function with λ determining the steepness of g. Following Hopfield (1984) we will let $\lambda = 1.4$. The R_j can be thought of as the electrical resistance of a neuron's plasma membrane, but for convenience we let $R_1 = R_2 = 1$.

When the inputs to a Hopfield network are held constant, they can be regarded as parameters of a dynamical system, which governs the network. In Hopfield (1984) the author shows that when the inputs to a Hopfield network are held constant, the network only has fixed point attractors. Hopfield's demonstration followed from his construction of a Lyapunov function for this network.[16] Knowing that there are only fixed point attractors implies that a stability analysis of the fixed points in any particular Hopfield network can explain the behavior of the network in response to fixed input values.

However, when the inputs vary over time, as they do here, a Hopfield network's state may not tend toward a fixed point. One way to see this is by imagining the Lyapunov function as varying in time in response to changing inputs.[17] This can be visualized by picturing a small ball rolling on a smooth surface. As long as the surface doesn't change its shape (which is the case when the network is not embedded in an environment), the ball will settle into a depression on the surface. But if the surface never ceases changing its shape (as occurs when the network is embedded in an environment), then the ball may never come to rest.

Hopfield networks are trained to recognize objects by using a Hebbian learning rule, which endows the network with a capacity for associative memory recall. We train our Hopfield network to be a two object detector, which enters the state $\xi_1 = (1,-1)$ when it

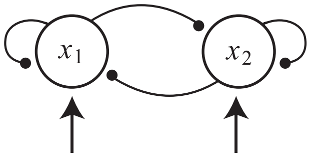

Fig. 7. Topology of the Hopfield network. Each node inhibits itself and the other node while receiving input from the environment.

detects object 1, and enters the state $\xi_2 = (-1,1)$ when it detects object 2. This determines the values for the weights $w_{12} = w_{21} = -1$.

The agent dynamical system, ϕ_α, is given by the solutions to Hopfield's differential equations

$$\dot{x}_1 = -x_1 - \frac{2}{\pi}\arctan(0.7\pi x_2)$$
$$\dot{x}_2 = -x_2 - \frac{2}{\pi}\arctan(0.7\pi x_1)$$

The phase portrait for ϕ_α is shown in Fig. 8. This system has two fixed point attractors, approximately (0.6,–0.6) and (–0.6,0.6), which occur near ξ_1 and ξ_2, respectively. We often assume that the state of the network begins at one of these fixed points. The diagonal line $x_1 = x_2$ is a separatrix between their basins of attraction: Initial conditions on either side of the diagonal line will approach the fixed point on that side.

The dynamics of the environment are the same as in the previous example; but for convenience, we define it in terms of the differential equation:

$$\dot{r}_1 = v_1$$
$$\dot{r}_2 = v_2$$

The environment's differential equation induces the environmental dynamical system $\phi_\epsilon(r_1,r_2,t) = (r_1,r_2)+t(v_1,v_2)$.

We define the open dynamical system as follows. The total state space is the Cartesian product of the state space for the environment and the state space for the neural network,

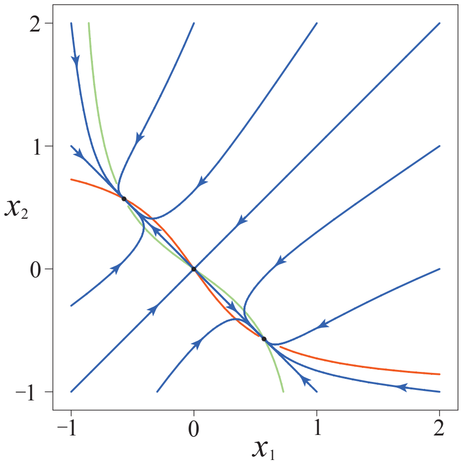

Fig. 8. A phase portrait for the decoupled Hopfield network, ϕ_α. The nullcline for $\dot{x}_1 = 0$ is shown in green and the nullcline for $\dot{x}_2 = 0$ is shown in red. The fixed points are located where the nullclines cross. Some of the other orbits are shown in blue. Compare these intrinsic dynamics, which involve decay to one of two attracting fixed points, with the behavior of the network when it is embedded in various environments.

that is, $S_\tau = \mathbf{T}^2 \times \mathbf{R}^2$. We let π be the projection of S_τ onto $\mathbf{R}^2$ and we let σ embed $\mathbf{R}^2$ as $\{(0°,0°)\} \times \mathbf{R}^2$. The total dynamical systems ϕ_τ and φ_τ are defined in terms of the parameterized family of differential equations

$$\begin{aligned}
\dot{r}_1 &= \mu v_1 \\
\dot{r}_2 &= \mu v_2 \\
\dot{x}_1 &= -x_1 - \frac{2}{\pi}\arctan(0.7\pi x_2) + \mu f(d(r_1)) \\
\dot{x}_2 &= -x_2 - \frac{2}{\pi}\arctan(0.7\pi x_1) + \mu f(d(r_2))
\end{aligned}$$

For each $\mu \in [0,1]$ the solutions to the differential equations determine a dynamical system

$$(r_1, r_2, x_1, x_2, t) \mapsto \Phi(r_1, r_2, x_1, x_2, t, \mu)$$

We set $\phi_\tau(r_1,r_2,x_1,x_2,t) = \Phi(r_1,r_2,x_1,x_2,t,1)$ and $\varphi_\tau(r_1,r_2,x_1,x_2,t) = \Phi(r_1,r_2,x_1,x_2,t,0)$. With $\mu = 0$ the angular position of the objects is fixed so $\sigma(S_\alpha) = \{(0°,0°)\} \times \mathbf{R}^2$ is invariant under φ_τ. Also, with $\mu = 0$, the inputs to the neural network are zeroed out so on $\sigma(S_\alpha)$, the dynamical system φ_τ is equivalent to ϕ_α. Thus, we have an open dynamical system model for this Hopfield neural network embedded in the circle world.

As noted above, in this example, it is difficult to identify all the isotypes and path topologies. However, we can consider the attracting paths, which occur in this system when the objects move at slow and fast velocities, and understand in a qualitative way how the resulting path topologies reflect different environmental configurations. At slow velocities, the distance between the objects dominates the representational process in the network (compare the panels in Fig. 9). At high velocities, the network behaves much differently, its state being confined to a very small region (see Fig. 12). In each case, we continue to make use of the indicator view of representation, whereby the states the network occupies when exposed to an object (or in the high velocity case, the two objects considered as a new kind of stimulus) correspond to representations of those objects. In both cases, this means that regions around the fixed point attractors of the closed system correspond to representations of one or both objects.

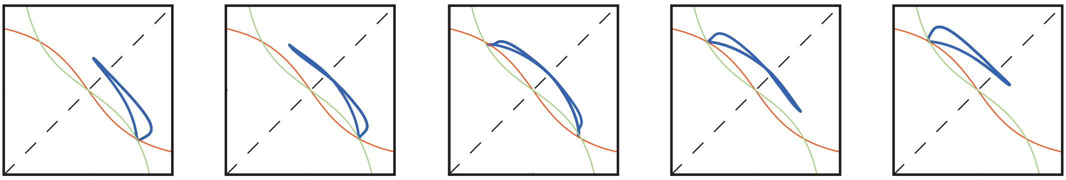

Fig. 9. Paths in the agent state space for the embodied Hopfield network when the objects move with speed 1/2. The paths are shown in blue, the nullclines are shown as red and green curves, and the $x_1 = x_2$ diagonal is shown as a dashed line. The five views show what happens when the angular separation between the objects is 60°, 90°, 180°, 270°, and 300°, respectively.

In the low-velocity case, the system follows the ‘‘fixed’’ points of the closed system, which move as the passing objects produce changing inputs to the network. The attracting paths of the agent system are approximated by the curves traced out by these slowly moving points. There are interesting differences between the sensory network and the Hopfield network. For instance, the specific value for the angular velocity of the slowly moving objects makes a difference to the shape of the paths in the Hopfield network while it does not for the sensory network (compare the paths in Fig. 6 with the paths in Fig. 9). This difference can be seen as a topological manifestation of the downstream processing that occurs when the outputs of the sensory network are fed to the Hopfield network (see note 14).

The behavior of the Hopfield network in response to particular environmental configurations is more complex than the behavior of the sensory network in response to the same configurations. We consider two specific configurations in more detail: the case of 180° angular separation, and the case of 60° angular separation.

In the case of 180° angular separation, the agent first represents one object, then the other, as they come in to and go out of view. This corresponds to the ‘‘L-shaped’’ path in the sensory network: the arc with two straight segments, where each segment corresponded to a representation of one object. The Hopfield network, by contrast, represents this configuration of passing objects with a figure eight–shaped curve (see Fig. 10). The outer portion of the figure eight is reached after an object has come in to view and the network has slowed down to a relatively steady state, which corresponds to representations of one of the objects (this captures the idea that only network states that are sufficiently stable produce normal sensory experience; see Koch, 2004, and Smolensky, 1988). The middle portion of the figure eight corresponds to transitional states between these representations.

The two loops of the figure eight–shaped path in the Hopfield network's state space are analogous to the two line segments of the ‘‘L-shaped’’ path in the sensory network's state space, insofar as each loop is associated with a representation of one of the objects.

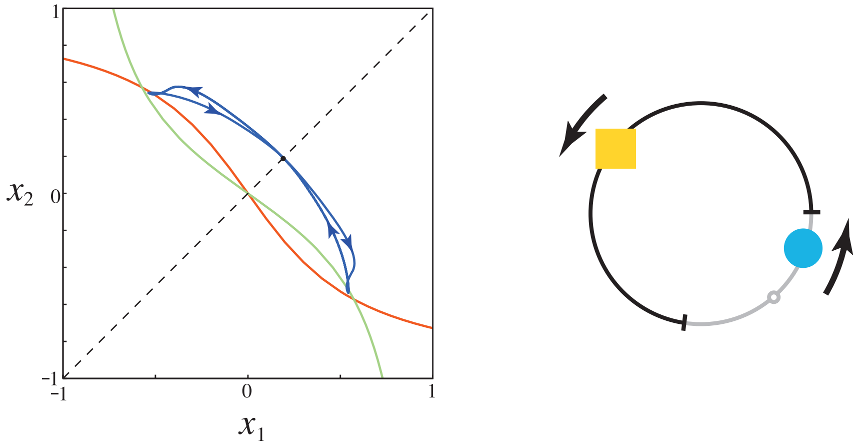

Fig. 10. The attracting path of the embedded Hopfield network when the objects move with speed 1/2 and are separated by 180°. Because of the wide separation between the objects, the network has time to produce a stable representation of each object.

However, the two loops also display hysteresis. The state of the sensory network goes back and forth over a line segment with the appearance and disappearance of an object, whereas for the Hopfield network, the state travels around a loop. The states followed by the Hopfield network as the object approaches the agent are different from the states followed as the object moves away from the agent. We almost always see such hysteresis effects in an embodied Hopfield network.[18]

Points in the middle of the figure eight path occur when an object first comes into view and the network is pulled out of one basin of attraction for the agent system, and into another. These network states may correspond to transitory, intermediate forms of representation. Such transient states have traditionally been overlooked in psychology because of (among other things) the discrete character of language. As Michael Spivey puts it, the discreteness of behavioral utterances ''is commonly misinterpreted as evidence for the internal discreteness of the mental representations that led to them'' (Spivey, 2006). Spivey goes on to describe experimental techniques for studying periods of ''fuzzy, graded, probabilistic mental activity,'' and in doing so, provide experimental support for a tradition that dates back at least to William James. In a famous passage, James (making explicit reference to what is in our terms ''path speed'') compares transitions between stable and transitional periods of mental life with the flights and perchings of a bird:

> When the rate [of change of mental and neural states] is slow we are aware of the object of our thought in a comparatively restful and stable way. When rapid, we are aware of a passage, a relation, a transition *from* it, or *between* it and something else … Like a bird's life, [the stream of thought] seems to be made of an alternation of flights and perchings.''
> (James, 1890, p. 243)

In the case of 60° angular separation with object 2 (the square) followed by object 1 (the disk), both objects can be in the agent's field of view at the same time. This case is represented by a loop-shaped path in the sensory network (see the bottom row of Fig. 6). The Hopfield network also represents this configuration of objects by a loop-shaped path (see Fig. 11). However, for the sensory network, both objects were represented by the network throughout the period in which they appeared. In the Hopfield network, by contrast, object 2 is never represented; it is ''masked'' by the appearance of object 1. When object 2 first appears, the network is momentarily pulled toward the representational region for object 2 (a neighborhood of the ''fixed point'' located above the separatrix). However, the network never reaches this representational region, since the subsequent appearance of object 1 draws the network back to the representational region for object 1. This is consistent with the phenomenon of perceptual masking (Breitmeyer & Ogmen, 2000), where a masking stimulus (in this case object 1) reduces an agent's representation of a target stimulus (in this case object 2).

When the two objects move sufficiently fast, a technique known as averaging can be used to approximate the behavior of the neural network. The idea is to determine the average value of the inputs I_1,I_2 to the two nodes of the neural network. This can be done by integrating the scaling function f over one period, T (the common period of r_1,r_2), and dividing by

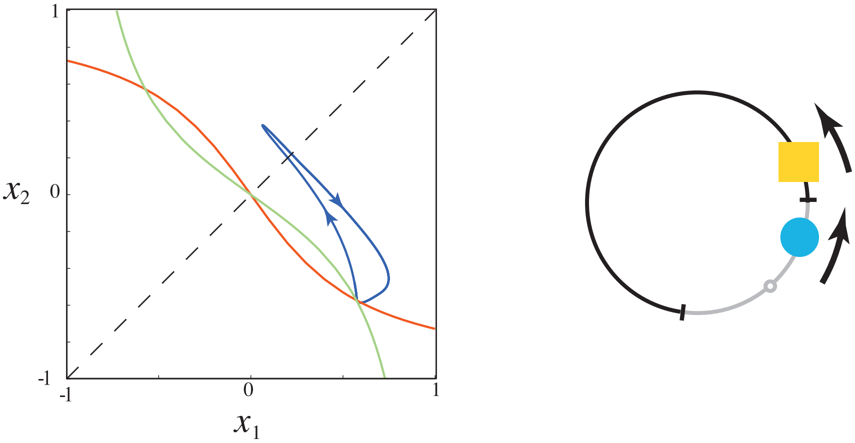

Fig. 11. The attracting path of the Hopfield network when the objects move with speed 1/2 and are separated by 60°. Note that the path does not enter the representational region for the square even though the agent is exposed to it. The disc ''masks'' the square so that only the disc is perceived.

the length of the period. When the speed is constant, the integral can be computed by invoking the formula for the area of an isosceles triangle. The ascending side of the triangle goes from 0 to 1 as an object comes in to view, the descending side goes from 1 to 0 as an object goes out of view, and the base has length 120°. So for $j = 1{,}2$

$$\bar{I}_j = \frac{1}{T}\int_0^T f(d(r_j(t)))dt = \frac{\frac{1}{2}\cdot 120\cdot 1}{360} = \frac{1}{6}$$

and thus the average input to each node in this case is 1/6. The differential equation for the averaged system is then

$$\dot{x}_1 = -x_1 - \frac{2}{\pi}\arctan(0.7\pi x_2) + \frac{1}{6}$$
$$\dot{x}_2 = -x_2 - \frac{2}{\pi}\arctan(0.7\pi x_1) + \frac{1}{6}$$

There are three fixed points in S_α for the averaged dynamical system. There is a saddle point on the diagonal $x_1 = x_2$ and two attracting fixed points symmetrically positioned about the diagonal. The Hopfield network's state travels to whichever attracting fixed point is closest to the initial condition (in this case, the angular separation between the objects has little impact on the path). Also note that averaging only provides an approximation of where the network state will go. The neural network's state actually follows one of two small tightly curved paths centered on the attracting fixed points of the averaged system. The larger the speed of the passing objects, the more tightly the paths wind around these attracting fixed points (see Fig. 12).

We can think of the high-velocity case as corresponding to a new type of stimulus, involving high-speed oscillating inputs to an agent's sensors. The attracting paths, which are near the fixed points of the averaged system, correspond to two distinct representations. Which representation occurs depends on which initial state the agent system began in. The behavior of the agent in this environment is consistent with the phenomenon of perceptual ambiguity or multistability, whereby ''more than one perceptual organization can be imposed on a stimulus'' (Hock et al., 1993, p. 63). Ambiguous figures (e.g., the Necker Cube, the face-vase, etc.) are subject to priming effects, whereby the interpretation of a stimulus is biased by a prior exposure to an unambiguous figure (Long, Toppino, & Mondin, 1992). In this example, we can think of the initial state of the agent system as corresponding to an initial biasing presentation by one of the objects, followed by a presentation of an ambiguous oscillating stimulus, which will be interpreted as either object 1 or object 2 depending on the initial presentation.

In addition to the analyses presented here, we have performed a number of additional experiments. We have added additional nodes to the simple sensory network and the Hopfield network. In the case of a three-node sensory network, the agent space contains 27 isotypes and nine types of exceptional paths (which reflect, in interesting ways, the additional types of perceptual process that can occur when a third object is present). We have considered the impact of learning (via backpropagation of error) on path shape in a three-layer network. When training a network to convert a distributed representation into a local representation, the appearance of the paths gradually changed from a loop into a collection of arcs attached at a point (a ''boquet of arcs'').[19] In the Hopfield case, we have analyzed the bifurcation that occurs when the object velocities increase. These bifurcations are consistent with perceptual phenomena like stroboscopic motion, where qualitative changes occur as the velocity of a set of stimuli is increased. We have also experimented with other velocities for the objects and in these cases more complex path shapes occur (e.g., Lissajous

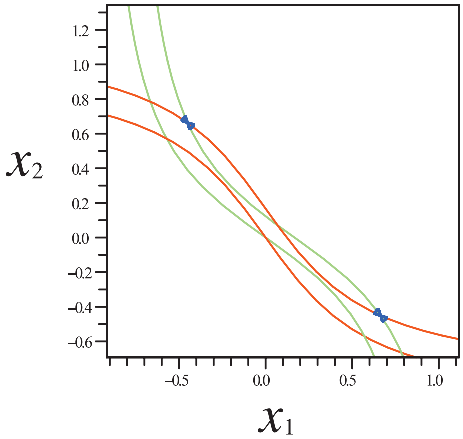

Fig. 12. The attracting paths for the Hopfield network in the diagonal world with a speed of $v_1 = v_2 = 10$. The nullclines for a system with inputs clamped at average values are also shown along with the nullclines for the closed agent system. The attracting paths are two small figure eights centered on the fixed points of the averaged system.

figures), and chaos is also observed (Hotton & Yoshimi, 2010). These results suggest that such analyses could be fruitful in interpreting representational processes in neural networks using the apparatus of open dynamical systems.

A third example of how open dynamical systems can be used to analyze the way an environment and agent act together to achieve a task comes from the analysis of development in a multicellular organism. The developmental pathways followed by the cells in a developing organism are largely determined by the cellular matrix the cells find themselves in. This allows specialized cells to appear in the proper position.

The cellular differentiation models of Furusawa and Kaneko (2002) and Yoshida, Furusawa, and Kaneko (2005) describe the rate of change for the concentration of chemicals inside of the cells with a set of differential equations. These concentrations tend to one set of values in the absence of external influences, but when the cells are coupled together, the concentrations shift from these baseline levels. We can therefore think of the internal chemical concentrations as representations in the simple sense of indication described above.

The solutions to the differential equations are dynamical systems for the chemical concentrations inside the cells. We can let the dynamical system for an isolated cell in their model be the agent system of an open dynamical system and the dynamical system for a collection of cells coupled together be the total system of an open dynamical system. There is one type of attractor (which they call a ‘‘single-cell attractor’’) for an isolated cell and another type of attractor (which they call a ‘‘partial attractor’’) for a cell coupled to other cells. A single-cell attractor corresponds to an attractor of an agent dynamical system; a partial attractor corresponds to an attracting path. States on the attracting paths for the ‘‘embedded’’ cell can function as a representation to the cell of its environment. These paths can be interpreted as representational process, which guides the activity of the developing cell. In this way, we can again see the value of open dynamical systems in the analysis of a representational process, in this case, a process guiding biological development.

In this section, we have considered what are sometimes called ‘‘idealized models’’ for the purpose of illustrating key concepts, in particular, neural net models with only two neurons and a simplistic mode of coupling to external objects. A model in this sense is ‘‘a deliberate simplification of something complicated with the objective of making it more tractable’’ (Frigg & Hartmann, 2006). Idealizations such as frictionless planes, point masses, and infinite velocities are often used in science. They leave a lot out, but they still capture important features of relevant phenomena and illustrate key principles in a perspicuous way. As Beer says in his discussion of embodied agents (which, as noted in Section 3, can be seen as open dynamical systems in our sense), ‘‘The early theoretical development of a field typically involves the careful study of simpler idealized models that capture the essential conceptual features of the phenomena of interest’’ (Beer, 2003). Although our illustrative examples are simple, our approach can be applied to more complicated cognitive processes. While more complicated systems can be more difficult to analyze, that does not diminish their value. Newtonian mechanics replaced a metaphorical view of the solar system with a precise mathematical model whose complexities are still being studied today. Despite the complexities of Newtonian mechanics, it remains the model of choice for computing orbital trajectories that send exploratory probes to destinations in the solar system. In our Hopfield example, we could not provide a complete classification of the isotypes, but we could still identify all the main types of attracting path and use them to obtain a robust understanding of the networks' way of representing its environment.

Moreover, many features of open dynamical systems scale quite naturally to more complex cases. For example, a complex embodied neural network with millions of neurons in a complex three-dimensional model environment could straightforwardly be interpreted as an open dynamical system with a total dynamics (the network/environment system), intrinsic agent dynamics (the neural network on its own), and distinctive open dynamics in each of its possible environments.

6. Reinterpretation of embodied cognition examples

In this section, we use the apparatus of open dynamical systems to reinterpret a range of key examples from the embodiment literature, to show how internalist and externalist ideas can be pursued in a common formal framework. In each case, we show how (1) embodiment is captured by a total system, and (2) important intrinsic dynamics are captured by one or more agent systems. In this way, (3) an agent's overall behavior can be understood as an alteration of its intrinsic dynamics relative to its ''embedding'' in an environment. This implies that traditional and embodied approaches to cognitive science can be pursued in a common framework. We also suggest (4) how (in at least some of these cases) we can usefully interpret paths as representational processes. Although we call this a ''reinterpretation,'' we will see that much of what we say is already implicit in many authors' work.[20]

A first, general class of examples (which subsumes most specific examples in the literature) involves cases where it is clear that the environment and agent work together in achieving some task, so that without the environment the task could not be completed. This is sometimes referred to as causal ''intimacy.'' As Haugeland says, ''The term intimacy is meant to suggest more than just necessary interrelation or interdependence but a kind of commingling or integralness of mind, body, and world—that is, to undermine their very distinctness'' (Haugeland, 1996, p. 208). Classic examples involve the behavior of insects and other simple agents, whose successful completion of tasks depends essentially on the way their bodies couple to environmental features. For example, Haugeland (1996) describes an ant's walk across a sand dune as involving not just its nervous system but also its legs and sensors and the shape of the dune itself. It is not so much that the ant knows how to get to its destination, but rather that the ant-body-dune system operates so as to result in the ant making it across the dune.

Haugeland is building on a long-standing tradition in ecological psychology, associated with Gibson (1986), who argued at length against the notion that the visual system is ''in the head,'' referring instead to an ''eye-head-brain-body-system'' (p. 61), and an ''animal-environment system'' (p. 225). A recent, theoretically oriented overview of this approach is Warren's ''The Dynamics of Perception and Action'' (Warren, 2006), which opens with the question ''How might one account for the organization in behavior without attributing it to an internal control structure?'' Warren proceeds to give a detailed analysis of several experiments, including pole balancing and obstacle avoidance, which can be understood in terms of dynamical systems incorporating agents and environments but no internal control structure.

Intimacy can be interpreted from the standpoint of open dynamical systems using points 1–4 above. (1) The total system of an open dynamical systems directly captures the kind of intimate coupling between agents and their environments emphasized by Haugeland, Gibson, Warren, and others. Just as the ''animal-environment'' system is basic for Gibson, the total system ϕ_τ is basic in this framework. Moreover, as explained in Section 3 above, dynamical systems form a special class of open dynamical systems. (2) However, this framework also emphasizes the importance of separately considering agent systems and their intrinsic dynamics (so, we do not follow Haugeland in insisting that the tight coupling of components undermines their distinctness).[21] Consider Haugeland's case of insect locomotion. In the literature on insect locomotion, central mechanisms—for example, central patten generators—have long played an important explanatory role (Wilson, 1966). These generators often have important intrinsic dynamics, for example, stable oscillatory dynamics. (3) By considering both intrinsic and environmental dynamics, traditional analyses of internal structures like central pattern generators can be pursued alongside embodied approaches. For example, we can understand the ant's behavior as an alteration of the generator's dynamics by the environment. In fact, this is the approach taken by Beer, a prominent exponent of dynamical and embodied views (see Chiel, Beer, & Gallagher, 1990).[22] Finally, (4), it seems plausible to suppose that in some cases of causal intimacy, paths in the relevant open system correspond to sequences of representations. For example, the changing state of an insect's sensors as it traverses a dune could correspond to its changing representation of the qualities of the dune surface, the pressure of the sand, or the visual scene.

A second, more specific class of examples involves cases where an agent ''offloads cognition,'' in the sense of using external artifacts to perform better at tasks than they otherwise would. For example, it has been shown (Hollan, Hutchins, & Kirsh, 2000; Maglio, Matlock, Raphaely, Chernicky, & Kirsh, 1999) that subjects can produce more words in the game *Scrabble* from a set of letters when they have the actual tiles before them and can physically manipulate them. In such cases, the *Scrabble* tiles themselves act as a working memory system for the agent. Other examples involve subjects using cell phones, calculators, the Internet, and other external artifacts to extend their cognitive powers (see Clark, 2008, sections 1.2–1.4, and chapter 4).

To capture this type of case in the present framework, we can (1) think of the total environment, including all embedded agents and artifacts (including both *Scrabble* players and their tiles), as our total system. (2) We can think of each *Scrabble* player + tiles combination as an agent system relative to this total system, and we can think of each player alone as an agent system relative to the corresponding *Scrabble* player + tiles combination (see Fig. 13). In this case, notice, we have a hierarchy of agent systems.[23, 24] Maintaining a focus on the subjects by themselves—without the tiles—allows us to think of each player as having intrinsic cognitive capabilities—for example, memory and search capacities—which are crucial to completing the overall task. This is consistent with how the original authors think of the example. For example, they say that subjects generate letter combinations ''in their heads,'' via an intrinsic ability to search a space of possible letter combinations for words. (3) When the subjects make use of the offloaded tile system, we can think of their intrinsic search capacity as being systematically (and beneficially) altered. This is how the original authors think of the example, describing the tiles as adding a kind of ''rubber band'' dynamic to intrinsic search: ''People can generate diverse letter combinations in their heads, but when they re-examine the tiles, they are drawn back to the original arrangement, like a rubber band springs back to its original shape'' (Maglio et al., 1999, p. 327). So, rather than thinking of the total system as a nondecomposable, holistic unit, the authors think of it as a collection of interacting subsystems, each making a distinct contribution. In this way, the traditional concept of memory retrieval as intrinsic search can be understood as making a specific, distinct contribution to the broader offloaded processes involving the external artifact. (4) Paths in some subspace of each player's set of brain states may correspond to representational processes, where various represented letter combinations are considered in a sequence whose shape reflects intrinsic search together with intermittent perceptual reanchoring.

A special case of offloading cognition involves functions normally performed by something internal to an agent being replaced by a functionally equivalent artifact. A famous example is the case of Otto and Inga (see Clark & Chalmers, 1998). Inga knows the location of a museum based on memories stored in her brain, whereas Otto has Alzheimer's and hence does not know the location of the museum, but he has written its location on a piece of paper. Since the message on the paper and the brain trace serve the same functional role, it is argued that we should treat the brain traces and paper notes as being equally cognitive.

Again, we can interpret this case using points 1–4 above. We can (1) think of the total system as the general environment that includes Otto and Inga as separate agents. (2) We can think of Inga as one agent system, and Otto with his notebook as a ''supplemented agent,'' while Otto himself, without his notebook, is a third agent system (this parallels the case of offloading cognition just discussed; see Fig. 14). This approach acknowledges that Otto has a significantly different cognitive system than Inga, insofar as the intrinsic

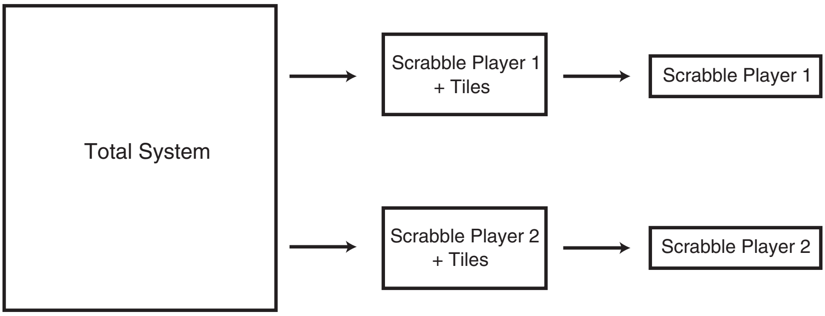

Fig. 13. A total system (corresponding to a game of *Scrabble*) and a hierarchy of subsystems, corresponding to *Scrabble* players with their tiles at the first tier, and *Scrabble* players without the tiles at the second tier.

dynamics of someone with Alzheimer's differs from that of a normal subject. (3) Given that Otto's intrinsic dynamics differ from Inga's, and that Otto's supplemented dynamics are similar to Inga's, it makes sense to suppose that the addition of the paper memory system amounts to an alteration of the structure of the paths in Otto's agent space so that they come to resemble the paths in Inga's agent space. This approach recommends a traditional analysis of what's missing with Alzheimer's, together with an analysis of how the external artifact supplements the Alzheimer's patient in a way that helps replace what's missing. (4) We can imagine that paths in a subspace of Otto's set of brain states and Inga's set of brain states correspond to the way they represent their environments over time. The shape of the paths in Otto's set of brain states may reflect the presence or absence of the memory aid, and again, we can hypothesize that there will be some structural similarity, at some level of abstraction, between the topology of the paths in supplemented Otto's set of brain states and those in Inga's.

A fourth class of examples involves cases where there is no single agent, but rather a larger system encompassing multiple agents and artifacts. Ed Hutchins (1995a; 1995b) has argued that such systems ''have interesting cognitive properties in their own right,'' referring to a ''wider dynamical process of which cognition of the individual is only a part'' (Hutchins, 1995a, p. xiii). Focusing on the case of a ship, Hutchins aims to ''move the boundaries of the individual person and treat the navigation team as a cognitive and computational system'' (Hutchins, 1995a, p. xiv). (1) The approach described here is sympathetic to the notion that the navigation team, and indeed the sea beyond and ultimately the whole universe, are dynamical systems (total systems) relative to which individual cognition is just one part. (2) This approach also advocates close attention to individual cognizers (and cognition/artifact combinations), which can be modeled as hierarchies of agent systems along the lines described above (e.g., in the discussion of *Scrabble*). Hutchins himself acknowledges the separate contributions of the various agents on the ship. For example, in the opening of Hutchins (1995a), Hutchins describes a placid scene on the U.S.S *Palau* as it was returning to port. The captain was ''watching the work of the bridge team'' (p. 1) the junior officer was ''directing the steering of the ship'' (p. 1) and the crew were talking about where to go for dinner. The ship subsequently experienced a loss of steam pressure, initiating a

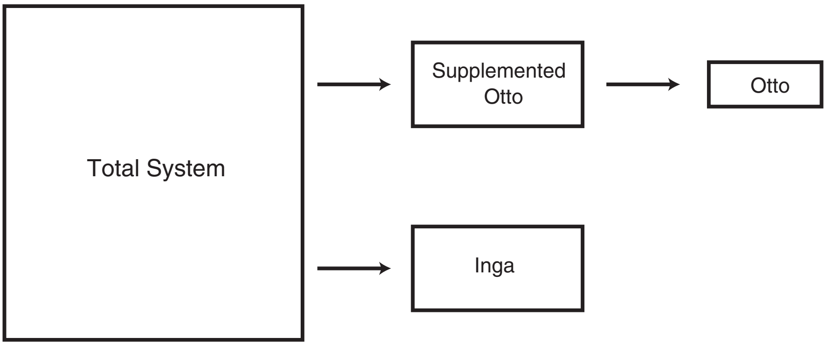

Fig. 14. Schematic of the Otto and Inga case.

frenzied series of events, including the fathometer operator ''reporting the depth of the water under the ship,'' the men on the cranks ''straining to move the rudder to the desired angle'' (p. 3), and the navigator remembering the location of a manual foghorn (p. 4). (3) The ship's behavior arises from coordinated interactions between these internal processes—these watchings, directings, reportings, strainings, and rememberings. Hutchins recognizes this. Referring to the crew's success in averting disaster, Hutchins says, ''Many kinds of thinking were required to perform the task … some inside the heads of individuals'' (pp. 5–6). In this way, traditional studies of perception, memory, attention, and so forth can be incorporated into an analysis of the ship as a cognitive system in its own right. (4) We can view paths in the agent space of a crew member's brain as representational processes reflecting that member's specific perspective on the ship as a whole.[25]

7. Conclusion

Embodied and dynamical system approaches to cognition have sometimes been presented as radical alternatives to traditional cognitive science, and as a means of overcoming the use of such traditional constructs as internal computational structures (e.g., central pattern generators) and representations, which mediate between sensory input and behavioral output. We think this rift is unwarranted. By extending dynamical systems theory as we have, traditional cognitive models can be incorporated into the study of embodied cognition. What we have in mind is more than a pluralistic reconciliation, where separate groups of researchers study internal processes and embodied processes independently. Rather, we have tried to show how the traditional approaches can be directly extended and thereby included in a unified framework.[26] For example, using these tools, one could begin with an ''internalist'' model of memory, learning, language, perception, or cognitive impairment, embed it in a plausible model environment, and study it using the tools of dynamical systems theory. This in turn makes it possible to compare a model's intrinsic dynamics with its open dynamics, identify isotypes and attracting paths in its open phase portrait (in some cases viewing these as representational processes), and understand how modifications of an agent's intrinsic dynamics are produced by a structured environment. As we have seen, this approach can be (and to some extent, already has been) used to study the way insects represent the places they explore, the different representational processes that occur in Alzheimer's patients that are deprived of and then given memory aids, and the way crew members understand the ships they are part of.[27]

Notes

1. This enterprise is associated with an extensive and somewhat inconsistent terminological apparatus. Approaches to cognitive science that emphasize the body and environment are variously referred to as ''embodied,'' ''embedded,'' ''extended,'' ''enactive,'' ''grounded,'' ''situated,'' ''ecological,'' and ''externalist.'' We stipulate the following usages. First, we take ''embodiment'' to refer broadly to

approaches to cognitive science that emphasize the importance of the body and environment. This can be a bit misleading, since ''embodied'' approaches are sometimes contrasted with ''extended'' approaches (where an embodied approach focuses on the role of the body and an extended approach focuses on the role of the environment). But the term has become common enough as a general designation that we use it that way here. Second, by ''externalism'' we mean the more specific philosophical claim that cognitive processes sometimes extend beyond the head (within the philosophical literature this type of externalism is sometimes referred to specifically as ''active externalism,'' to distinguish it from several other forms of philosophical externalism). Finally, there are some who use the term ''embodiment'' to refer to the way abstract concepts are grounded in bodily schemata, and are thus doubtful about the possibility of purely conceptual thought (Barsalou, 2008; Lakoff & Johnson, 1999). This is also referred to as ''grounded cognition.'' We do not address these issues here.

[2] Others have described similar approaches to dynamical systems theory, for example, Beer in van Gelder and Port (1995), Petitot in van Gelder and Port (1995), Carello in van Gelder and Port (1995). Also see Chemero (2009), Kelso (2001), Schmidt, Carello, and Turvey (1990), Thelen and Smith (2006), and Warren (2006). We elaborate on the relationship between our view and Beer's and Warren's below.

[3] The framework we present here is distinct from several related ideas with similar nomenclature, in particular in control theory and thermodynamics. For more on the relationship between our view and the control theory case, see Hotton and Yoshimi (2010). The thermodynamics case is especially relevant because some in the embodied cognition/dynamical systems literature cite this usage, for example, Schmidt et al. (1990) and Thelen and Smith (2006). In open thermodynamic systems, energy can flow in to and out of a system. However, the emphasis is on relatively simple forms of interaction with outside energy sources, and on states near equilibrium. By contrast, we allow environments to have complicated dynamics and focus on the behaviors of agents that are embedded in such environments.

[4] A more detailed introduction to dynamical systems theory for psychologists can be found in Abraham, Shaw, and Abraham (1990), and for neuroscientists in Izhikevich (2007).

[5] The time space is often taken to be the set of real numbers, but other sets can be used for the time space such as the set of integers or even just the set of nonnegative integers. The use of a discrete set for time is usually a matter of convenience and is not meant to imply that time itself is fundamentally discrete. In this article we let the time space be the real numbers.

[6] To see this note that the state s_0 must go to the state $\phi(s_0,t_1)$ at time t_1. The state $\phi(s_0,t_1)$ can then be taken as an initial condition and it must go to the state $\phi(\phi(s_0,t_1),t_2)$ at time t_2. If the state s_0 did not end up at the state $\phi(\phi(s_0,t_1),t_2)$ at time $t_1 + t_2$, then the future of state s_0 would not be uniquely determined.

[7] Differential equations and dynamical systems are, strictly speaking, distinct concepts. It is the general solution to an ordinary differential equation that satisfies the existence and uniqueness theorem which can be a dynamical system.

8. Thus, our approach is more general than the dynamical systems approach to cognitive science (the ''dynamical hypothesis'' in cognitive science; van Gelder, 1998, and van Gelder and Port, 1995), which is presented as an alternative to connectionist and symbolic AI approaches. Advocates of this approach emphasize a specific style of cognitive modeling, which precludes the use of internal representations, an integer time set, discrete state variables, etc. We do not preclude any of these.
9. More precisely, the set of all future states under a dynamical system starting from some initial state is called the *forward orbit* of the initial state. The set of all states whose forward orbits contain the initial state is called the *backward orbit* of the initial state. The union of the forward and backward orbits of a state is called the *orbit* of the state.
10. The name ''isotype'' has been used in Golubitsky, Schaeffer, and Stewart (1988) to classify the orbits of a dynamical system in terms of their symmetry properties. In some instances the two uses of the term may agree, but generally they are distinct concepts.
11. We also assume that σ is compatible with the projection: if an agent state $x \in S_\alpha$ is taken to total state $y \in S_\tau$ under the map σ, then the projection π takes y back to x. That is σ is a right-inverse of π, $\pi \circ \sigma = \mathrm{id}_{S_\alpha}$.
12. For the examples presented here ''equivalence'' of dynamical systems will mean ''topological equivalence.'' Formally two dynamical systems, $\phi_1 : S_1 \times \mathbf{R} \to S_1$, $\phi_2 : S_2 \times \mathbf{R} \to S_2$, are *topologically equivalent* if there is a homeomorphism $f : S_1 \to S_2$, and a continuous map $g : S_1 \times \mathbf{R} \to \mathbf{R}$ such that for each $s \in S_1$ the map $t \mapsto g(s,t)$ is an orientation preserving homeomorphism of $\mathbf{R}$ and such that $f(\phi_1(s,g(s,t))) = \phi_2(f(s),t)$ for all $s \in S_1$ and $t \in \mathbf{R}$. The homeomorphism f maps orbits in S_1 to orbits in S_2 and the map g reparameterizes time along the orbits.
13. In fact, we could allow the total space and agent space to be rather exotic; for instance, S_τ could be a Möbius strip or a Klein bottle, S_α could be a circle, and the map π could be a fiber bundle projection. Though we are not aware of examples in cognitive science where the total state space would have these forms, we do not preclude their possibility. However, there is an interesting example in the field of image analysis of Klein bottles arising in the space of 3×3 patches of digitized images (Carlsson, Ishkhanov, & de Silva, 2008).
14. We can think of the two examples as describing the behaviors of two ''layers'' in a model system: first, the behavior of a set of sensory units that are coupled directly to an environment, and second, the behavior of a set of downstream, recurrently connected processing neurons. From that perspective, we analyze representational processes in two subsystems of a single agent: a sensory system, and a more abstract conceptual system.
15. Indeed, authors who make use of embodiment and dynamical systems theory often refer to ''shifting attractor landscapes,'' ''movement between basins of attraction,'' etc. See, for example, Smolensky (1988), Spivey (2006), Thelen and Smith (2006), and van Gelder and Port (1995). The discussion in Thelen and Smith places the dis-

cussion in historical context (referring, for example, to Waddingtons' work on epigenetic landscapes, which he also referred to as ''attractor landscapes''). As noted above, one of our purposes here is to put these types of discussion in the context of a rigorous mathematical framework.

16. Lyapunov functions are a class of functions designed to show the stability of fixed points.
17. By definition Lyapunov functions do not vary with time, so technically it would be better to say that there ceases to be a Lyapunov function for a Hopfield network when the inputs vary.
18. The stretched out loops of the figure eight path may not resemble some of the more familiar forms for hysteresis loops, and so might not be immediately recognized as such (see Hock, Kelso, & Schöner, 1993). However, we can clearly see the key property of hysteresis in this case, that is, the same stimulus produces different agent states, depending on the recent history of the embodied agent. For more on the application of open dynamical systems theory to the concept of hysteresis, see Hotton and Yoshimi (2010).
19. Strictly speaking, the topology of the path did not change, but its shape changed so that it very closely approximated a bouquet of arcs.
20. A note regarding the structure of our argument. Throughout this section, we assume that there are sound arguments on both sides of the embodiment debate: that is, sound arguments that conclude that internal representations, intrinsic dynamics, and other internal structures are important in cognitive research; and sound arguments that conclude that in some cases, it is useful to think of cognitive processes as extending to an agent's body and environment. Our intention is to show how such arguments can be mutually consistent (even though they are sometimes presented as arguments against one another), and even mutually supporting. For example, when we claim that central pattern generators (CPGs) are an important explanatory construct, we are accepting that others have argued this point successfully (we are ''offloading'' that work). Our primary goal is to show how this idea can be fruitfully pursued alongside embodied approaches to insect behavior.
21. Insofar as we emphasize the distinct contributions made by the intrinsic dynamics of an agent system, we follow Adams, Aizawa, and Rupert, among others (see Adams & Aizawa, 2001; Rupert, 2004; and Clark, 2008, chapters 5–6).
22. It is also to Warren's credit that he, at least tacitly, recognizes these points, insofar as he separately defines agent and environment dynamical systems, as we do, and thereby facilitates separate analysis of each on its own followed by analysis of the two together.
23. Cashing out how such hierarchies work in terms of the formal framework developed above turns out to be a nontrivial task. However, the details are mostly technical, and the idea of a subsystems hierarchy is straightforward enough to be discussed in a nonformal way here.
24. Compare the images Beer uses (in several places) to describe brain-body-world interactions. See, for example, Beer (2000), p. 97.

25. Compare Grush's discussion of a pilot operating an airplane (Grush, 2003, p. 81).
26. The pluralistic approach is discussed in Dale, Dietrich, and Chemero (2009) and a special issue of the *Journal of Experimental & Theoretical Artificial Intelligence*, 20(3), September 2008, devoted to ''Pluralism and the Future of Cognitive Science.'' In his article in that issue, Dale (2008) encourages cognitive researchers to ''embrace diversity'' (as we do), but he is doubtful about efforts to find a ''single overarching theory or framework'' for cognitive science (also see Dale et al., 2009; Dietrich, 2008). While this caution is warranted, we think that it is sometimes possible, using tools like the ones we have described, to integrate systematically different approaches in a single framework (see Jilk, Lebier, O'Reilly, & Anderson, 2008; Markman, 2008; Spivey & Anderson, 2008).
27. We are grateful to Anthony Chemero, Chris Kello, Michael Spivey, the editor, and the referees for helpful comments.

References

Abraham, F., Shaw, C., & Abraham, R. (1990). *A visual introduction to dynamical systems theory for psychology*. Santa Cruz, CA: Aerial Press.

Adams, F., & Aizawa, K. (2001). The bounds of cognition. *Philosophical Psychology*, *14*(1), 43–64.

Barsalou, L. (2008) Grounded cognition. *Annual Reviews of Psychology*, *29*, 617–645.

Beer, R. (2000). Dynamical approaches to cognitive science. *Trends in Cognitive Sciences*, *4*(3), 91–99.

Beer, R. (2003). The dynamics of active categorical perception in an evolved model agent. *Adaptive Behavior*, *11*(4), 209–243.

Breitmeyer, B., & Ogmen, H. (2000). Recent models and findings in visual backward masking: A comparison, review, and update. *Perception & Psychophysics*, *62*(8), 1572–1595.

Carlsson, G., Ishkhanov, T. & de Silva, V., (2008). On the local behavior of spaces of natural images. *Internation Journal of Computer Vision*, *76*, 1–12.

Chemero, A. (2009). *Radical embodied cognition science*. Cambridge, MA: MIT Press.

Chiel, H., Beer, R., & Gallagher, H. (1990). Evolution and analysis of model CPGs for walking: I. Dynamical modules. *Journal of Computational Neuroscience*, *7*, 99–118.

Churchland, P., & Sejnowski, T. (1992). *The computational brain*. Cambridge, MA: MIT Press.

Clark, A. (2008). *Supersizing the mind: Embodiment, action, and cognitive extension*. Oxford, England: Oxford University Press.

Clark, A. (2009). Spreading the joy? why the machinery of consciousness is (probably) still in the head. *Mind*, *118*, 963–993.

Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis*, *58*(1), 7–19.

Dale, R. (2008). The possibility of a pluralist cognitive science. *Journal of Experimental & Theoretical Artificial Intelligence*, *20*(3), 155–179.

Dale, R., Dietrich, E., & Chemero, A. (2009). Explanatory pluralism in cognitive science. *Cognitive Science*, *33*(5), 739–742.

Dietrich, E. (2008). Pluralism, radical pluralism and the politics of the Big Bang. *Journal of Experimental & Theoretical Artificial Intelligence*, *20*(3), 231–237.

Frigg, R., & Hartmann, S. (2006). Models in Science, In E. Zalta (Ed.), *The Stanford encyclopedia of philosophy (Summer 2009 edition)*. Stanford, CA: Metaphysics Research Lab, CSLI. Available at: http://plato.stanford.edu/archives/sum2009/entries/models-science/. Accessed on July 23, 2010.

Furusawa, C., & Kaneko, K. (2002). Origin of multicellular organisms as an inevitable consequence of dynamical systems. *The Anatomical Record*, *268*, 327–342.

van Gelder, T., (1998). The dynamical hypothesis in cognitive science. *Behavioral and Brain Science*, *21*, 615–665.

van Gelder, T., & Port, R. (1995). *Mind as motion: Explorations in the dynamics of cognition*. Cambridge, MA: MIT Press.

Gibbs, R. (2006). *Embodiment and cognitive science*. Cambridge, UK: Cambridge University Press.

Gibson, J. (1986). *The ecological approach to visual perception*. New York: Taylor and Francis.

Golubitsky, M., Schaeffer, D., & Stewart, I. A. (1988). *Singularities and groups in bifurcation theory: Volume 2 (applied mathematical sciences)*. New York: Springer.

Grush R. (2003). In defense of some 'Cartesian' assumptions concerning the brain and its operation. *Biology and Philosophy*, *18*, 53–93.

Hasselblatt, B., & Katok, A. (2002). *Handbook of dynamical systems, volume 1A*. Amsterdam: North-Holland Publishing Company.

Haugeland, J. (1996). Mind embodied and embedded. In J. Haugeland (Ed.), *Having thought: Essays in the metaphysics of mind* (pp. 207–237). Cambridge, MA: Harvard University Press.

Hock, H., Kelso, S., & Schöner, G. (1993). Bistability and hysteresis in the organization of apparent motion patterns. *Journal of Experimental Psychology*, *19*(1), 63–80.

Hollan, J., Hutchins E., & Kirsh, D. (2000). Distributed cognition: Toward a new foundation for human-computer interaction research. *ACM Transactions on Computer-Human Interaction*, *7*(2) 174–196.

Hopfield, J. (1984). Neurons with graded response have collective computational properties like those of two-state neurons. *Proceedings of the National Academy of Science*, *81*(10), 3088–3092.

Hotton, S., & Yoshimi, J. (2010). The dynamics of embodied cognition. *International Journal of Bifurcations and Chaos*, *20*(4), 943–972.

Hutchins, E. (1995a). *Cognition in the wild*. Cambridge, MA: MIT Press.

Hutchins, E. (1995b). How a cockpit remembers its speeds. *Cognitive Science*, *19*, 265–288.

Izhikevich, E. (2007). *Dynamical systems in neuroscience: The geometry of excitability and bursting*. Cambridge, MA: MIT Press.

James, W. (1890). *The principles of psychology*. New York: Holt. Available at: http://psychclassics.yorku.ca/James/Principles/index.htm. Accessed on July 23, 2010.

Jilk, D., Lebier, C., O'Reilly, R. & Anderson, J. (2008). SAL: An explicitly pluralistic cognitive architecture. *Journal of Experimental & Theoretical Artificial Intelligence*, *20*(3), 197–218.

Kelso, J. A. S. (2001). *Dynamic patterns: The self-organization of brain and behavior*. Cambridge, MA: MIT Press.

Koch, C. (2004). *The quest for consciousness*. Englewood, CO: Roberts and Company.

Lakoff, G. & Johnson, M. (1999). *Philosophy in the flesh: The embodied mind and its challenge to Western thought*. New York: Basic Books.

Long, G., Toppino, T., & Mondin G., (1992). Prime time: Fatigue and set effects in the perception of reversible figures. *Perception and Psychophysics*, *52*(6), 609–616.

Maglio, P., Matlock, T., Raphaely, D., Chernicky, B., & Kirsh, D. (1999). Interactive skill in scrabble. In M. Hahn & S. C. Stoness (Eds.), *Proceedings of twenty-first annual conference of the Cognitive Science Society* (pp. 326–330). Mahwah, NJ: Lawrence Erlbaum Associates.

Markman, A. (2008). Pluralism, relativism and the proper use of theories. *Journal of Experimental & Theoretical Artificial Intelligence*, *20*(3), 247–250.

Markman, A., & Dietrich, E. (2000a). Extending the classical view of representation. *Trends in Cognitive Science*, *4*(12), 470–475.

Markman, A., & Dietrich, E. (2000b). In defense of representation. *Cognitive Psychology*, *40*, 138–171.

Neander, K. (2004). Teleological theories of mental content. In E. Zalta (Ed.), *Stanford encyclopedia of philosophy (Winter 2009 edition)*. Stanford, CA: Metaphysics Research Lab, CSLI. Available at: http://plato.stanford.edu/archives/win2009/entries/content-teleological. Accessed on July 23, 2010.

Noë, A. (2004). *Action in perception*. Cambridge, MA: MIT Press.

Robinson, C. (1995). *Dynamical systems: Stability, symbolic dynamics, and chaos*. Boca Raton, FL: CRC Press.

Rupert, R. (2004). Challenges to the hypothesis of extended cognition. *Journal of Philosophy*, *101*(9), 389–428.

Schmidt, R. C., Carello, C., & Turvey, M. T. (1990). Phase transitions and critical fluctuations in the visual coordination of rhythmic movements between people. *Journal of Experimental Psychology: Human Perception and Performance*, *16*, 227–247.

Smolensky, P. (1988). On the proper treatment of connectionism. *Behavioral and Brain Sciences*, *11*, 1–74.

Spivey, M. (2006). *The continuity of mind*. Oxford, England: Oxford University Press.

Spivey, M., & Anderson, S. (2008). On a compatibility between emergentism and reductionism. *Journal of Experimental & Theoretical Artificial Intelligence*, *20*(3), 239–245.

Thelen, E. & Smith, L. (2006). Dynamic systems theories. In W. Damon (Ed.), *Handbook of child psychology, volume 1, theoretical models of human development* (pp. 258–312). Hoboken, NJ: John Wiley & Sons.

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

Wilson, D. (1966). Insect walking. *Annual Reviews of Entomology*, *11*, 103–122.

Yoshida, H., Furusawa, C., & Kaneko K. (2005). Selection of initial conditions for recursive production of multicellular organisms. *Journal of Theoretical Biology*, *233*(4), 501–514.

Appendix: Open analog of an attractor

There are slight variations in the dynamical systems theory literature on the definitions of an attracting set and of an attractor, but most of these distinctions are not relevant here (e.g., are such sets attracting on an open set or only on a set of positive measure? See Hasselblatt & Katok, 2002; Robinson, 1995). For example, if we assume S is a metric space, we could define attractors in terms of convergence to a set. However, to remain as general as possible, we use a very broad, topological definition: A subset $A \subset S$ of the state space of a dynamical system $\phi : S \times T \rightarrow S$ is an *attracting set* if it is closed, strictly invariant, and there exist open sets U,V with $A \subset V \subset U$ and with the property that for all $s \in U$ there exists $t_s \in T$ such that the forward orbit of $\phi(s,t_s)$ is contained in V (''$\subset$'' refers here to proper containment). Any orbit that enters U is a basin orbit and can be thought of as ''trapped'' by V, since once it enters V it never leaves V. An *attractor* is an attracting set that contains a dense orbit.

The analog in open dynamical systems for an orbit of an attracting set is an attracting path. Suppose we have an attracting set, $A \subset S_\tau$, of the total system, ϕ_τ, and open sets U,V with $A \subset V \subset U$ satisfying the condition that for each $s \in U$ there exists $t_s \in T$ such that the forward orbit of $\phi_\tau(s,t_s)$ is contained in V. The projection of an orbit in A under π is an attracting path. The open analog of a basin orbit is a *trapped path*, and it is simply the projection of a basin orbit in the total space: The projection of an orbit that intersects U in the total space and is thereby trapped in V is a trapped path.