THE DYNAMICS OF EMBODIED COGNITION*

Scott Hotton [a], Jeff Yoshimi [b]

[a] *Department of Organismic and Evolutionary Biology, Harvard University, Cambridge MA 02138*

[b] *School of Social Sciences, Humanities, and Arts and UC Merced Center for Computational Biology University of California at Merced, P.O. Box 2039, Merced CA 95344*

Abstract

Historically cognition was understood as the result of processes occurring solely in the brain. Recently, however, cognitive scientists and philosophers studying "embodied" or "situated" cognition have begun emphasizing the role of the body and environment in which brains are situated *i.e.* they view the brain as an "open system". However, these theorists frequently rely on dynamical systems which are traditionally viewed as closed systems. We address this tension by extending the framework of dynamical systems theory. We show how structures which appear in the state space of an embodied agent differ from those that appear in closed systems, and we show how these structures can be used to model representational processes in embodied agents. We focus on neural networks as models of embodied cognition.

Keywords: connectionism, dynamical systems, embodiment, Hopfield networks, neural networks, open systems, situated cognition

Contents

*Penultimate draft of https://www.worldscientific.com/doi/abs/10.1142/S0218127410026241

1 Introduction

Many forms of cognitive activity cannot be understood unless one takes account both of the internal dynamics of an agent and its interactions with an environment. Long multiplication (Rumelhart *et al.* [1986]), ship navigation (Hutchins [1995]), coordinated behavior (*e.g.* walking and predation, Chiel & Beer [1997]), *Scrabble* (Kirsch [1998]), and the video game *Tetris* (Kirsch & Maglio [1994]) are among the examples that have been studied.[1] Many in the "embodied cognition" tradition, which studies these types of coupling, appeal to dynamical systems theory to capture the kinds of fluid interplay between agent and environment they are interested in.[2] But this approach is limited by the fact that, in order to ensure unique futures, dynamical models of an agent must either assume an agent is cut off from its environment or assume it is subject to a very simple environment.[3] Those who have attempted to incorporate more complex environments have been forced to use key terminology in an inconsistent way. For example fixed points are allowed to "move", though such points are obviously no longer "fixed". Trajectories are allowed to "change course" and follow different attractors as inputs change, though trajectories are by definition deterministic. Orbits are allowed to intersect themselves, even though such an object is no longer, strictly speaking, an orbit. Parameters are allowed to slowly vary even though parameters by definition are fixed. The very concept of hysteresis, as we shall see, does not have a precise mathematical definition. In each case authors are expressing ideas relating to systems which are in some sense "open" while working within the present framework of closed systems.[4,5] We will see that by formally defining the concept of an "open dynamical system" we can retain some of these ways of talking, *e.g.* in terms of moving fixed points and hysteresis loops, but we can do so against the background of an understanding of the relevant mathemati-

[1] The most extreme proponents of this perspective have gone so far as to argue that cognitive systems actually extend beyond the skin: " the human organism is linked with an external entity in a two-way interaction, creating a *coupled system* that can be seen as a cognitive system in its own right. If we remove the external component the system's behavioral competence will drop, just as it would if we removed part of its brain. Our thesis is that this sort of coupled process counts equally well as a cognitive process, whether or not it is wholly in the head" (Clark & Chalmers [1998]). Although we do not follow the authors in their attempt to overthrow the "hegemony of skin and skull," we share their belief that it is essential to study couplings between organisms and their environments. In particular, we assume that representation do occur "inside an agent," but believe that the order in which these representations occur depends on the agent's interaction with an environment.

[2] This is most explicit in Clark [1997], which opens with the "Radical Embodied Cognition Thesis," according to which "Embodied cognition is best studied using... explanatory schemes involving, *e.g.* dynamical systems theory" (Clark [1997], p. 461; also see pp. 465, 470, 475-476). The approach is attributed to a wide range of thinkers, most of whom are represented in the anthology by van Gelder & Port [1995]. Also see van Gelder [1998], where a key component of the "dynamical hypothesis in cognitive science" is the ability of dynamical systems theory to better handle agent/environment interactions, which involve "two systems simultaneously shaping each other's change" (p. 622).

[3] For example, model neurons often incorporate a simple input term, which is either clamped at a constant value or given a simple functional form (*e.g.* a sine wave).

[4] Some examples: In his article on language as a dynamical system (chapter 8 of van Gelder & Port [1995]), Jeffrey Elman describes the human lexicon as a structured state space which is organized into regions associated with different words; the regions the network visits as it is exposed to successive words in a sentence is described as a "trajectory," even though inputs and hence, the positions of attractors, are changing during sentence presentation (p. 210). Port, Cummins, and McAuley (chapter 12 of van Gelder & Port [1995]) also refer to trajectories changing course as inputs change: "Once the input changes to B, the attractor layout also changes and the system trajectory changes course toward the new attractor" (p. 358). Also see the "trajectories" plotted in Thelen's article on pp. 88-89 of van Gelder & Port [1995], which did not arise from fixed inputs but rather from continuously varying inputs.

[5] Some in the embodied cognition/dynamical systems field are clearly sensitive to these issues, *e.g.* Petitot in van Gelder & Port [1995] p. 237, Turvey and Carello in van Gelder & Port [1995] p. 375, Warren [2006], Schmidt *et al.* [1990] (especially p. 228), and Beer in van Gelder & Port [1995]. Beer, for example, describes an agent and its environment as coupled dynamical systems, and notes "each of these two dynamical systems is continuously deforming the flow of other ... and therefore influencing its subsequent trajectory" (p. 131) though even he is forced to use the term "trajectory" in a non-traditional way.

cal objects. In this way dynamical systems theory can be rigorously applied to the study of embodied cognitive agents.

We define open dynamical systems as a kind of compound system in which one system (corresponding to the agent by itself) is embedded in another (corresponding to the agent in an environment). These systems go through sequences of states ("paths") which display topological properties impossible in a classical dynamical system, for example, figure eights, bouquets of loops, and Lissajous figures. These can be interpreted as the sequences of representations that occur when an agent is open to an environment.

In section 2 we rigorously define open dynamical systems and some basic tools we use to analyze them. We use these tools to show how certain phenomena (*e.g.* hysteresis) can be more rigorously analyzed. In section 3 we give an overview of neural networks.[6] We show how various features of neural network models can be understood from an open dynamical systems perspective. In section 4 we embed a continuous Hopfield network in a simple environment to study how the intrinsic dynamics of an agent change when it is coupled to an environment. We also show how known perceptual phenomena such as masking and perceptual ambiguity emerge quite naturally, in this example, from the interaction of an agent's intrinsic dynamics with those of an environment.[7]

[6]Some in cognitive science who rely on dynamical systems theory are critical of neural network theorists insofar as, among other things, they "have not been utilizing dynamical concepts and tools to any significant degree" (van Gelder & Port [1995], p. 32). However many neural networks theorists do make use of dynamical systems theory (*e.g.* Hopfield [1984], who we reference in the final section) and, as we will try to show, neural networks are often amenable to analysis as dynamical systems.

[7]The concept of an "open dynamical system" brings to mind the concept of an "open system" in control theory, but the two concepts are distinct. In control theory an open system has a state variable and a control parameter which is free to vary continuously however one chooses. The goal of open systems theory is to see if the control parameter can be used to send the state variable to a desired value. Generally there is no dynamical model for the environment. In contrast, we consider an agent and its environment as a single system and introduce techniques to focus on what occurs inside the agent as it interacts with its environment. The concept of an "open system" is also used in the thermodynamics (and some in the embodied cognition / dynamical systems literature cite this usage of the concept, for example Schmidt *et al.* [1990]). In such systems energy is allowed to flow in to and out of a system. In this sense, open thermodynamic systems are conceptually similar to the systems we consider. However, in the thermodynamics context the emphasis is on relatively simple forms of interaction with outside energy sources, and mathematically, the emphasis is on solving for equilibrium states. Our emphasis is more directly on the concept of openness as it can be used in cognitive science. In particular, we consider environments that can have arbitrarily complex dynamics, and we focus not on solving for equilibrium states, but on understanding the complex behaviors that occur inside an agent when it is embedded in such an environment.

2 Mathematical preliminaries

In this section we describe a formal framework for analyzing embodied agents. In 2.1 we give some background on dynamical systems theory. In 2.2 we show how a dynamical system open to an environment differs from a classical dynamical system, and we show how this is relevant to modeling embodied cognition. In 2.3 we define some concepts that are useful in analyzing the behavior of an embodied agent. Finally, in 2.4, we use these tools to analyze three classical examples of dynamical systems, to illustrate the meaning and value of these concepts.

2.1 The dynamical systems framework

For the purpose of illustrating how we can rigorously apply dynamical systems theory to issues raised by embodied cognition we will confine ourselves in this tutorial to a fairly simple class of dynamical systems. Although it is not the most general form for dynamical systems studied in mathematics (*e.g.* Robinson [1995]; Hasselblatt & Katok [2002]) it is sufficient for illustrating the basic logical structure of dynamical systems and how one can go about using dynamical systems theory with embodied cognition.

A dynamical system consists of a state space, S, a time space, in our case the real numbers $\mathbf{R}$, and a continuous map

$$\phi : S \times \mathbf{R} \to S$$

that takes a state $s_0 \in S$ (which we think of as an initial condition) and a time $t \in \mathbf{R}$ and returns the state the system will be in at time t starting from state s_0.

We restrict attention to cases where the state space is a Cartesian product of some number (possibly zero) of copies of the real line, $\mathbf{R}$, and some number (possibly zero) of copies of the circle, $\mathbf{S}^1$. A state will be represented by an n-tuple of real numbers. Some components of the n-tuple may stand for points in a line and some components may stand for points on a circle. The exact form of the state space will be clearly expressed in each example. For technical reasons we also let a set with exactly one element be a state space. This state space will be denoted by $\{0\}$. Otherwise a state space will always have at least one copy of $\mathbf{R}$ or at least one copy of $\mathbf{S}^1$ in its Cartesian product. The dimension of the state space is called the dimension of the dynamical system. In separate work we consider the more general case of dynamical systems with arbitrary state spaces and time spaces (subject to the conditions below).

To be a dynamical system the map ϕ must satisfy two properties:

For all $s_0 \in S$

$$\phi(s_0, 0) = s_0 \tag{1}$$

For all $s_0 \in S$ and all $t_1, t_2 \in \mathbf{R}$

$$\phi(s_0, t_1 + t_2) = \phi(\phi(s_0, t_1), t_2) \tag{2}$$

The first property essentially means that we are letting 0 stand for the initial moment in time. The second property means that we can find where state s_0 goes at time $t_1 + t_2$ by first finding where s_0 goes at time t_1, treating that state as an initial condition, and finding where it goes from that initial condition at time t_2. In other words there are not multiple routes to the future. The future is uniquely determined by the state of the system in the present moment.

One consequence of our restriction to dynamical systems whose time space is the reals is that all the dynamical systems we consider have a property known as invertibility, which roughly means that it is possible to go backwards in time. If we know what state an invertible dynamical system is in at any particular moment then we can find the initial state of the system. To see this suppose that the state of an invertible system is s_1 at time t_1. Then one possible state that could have been the initial condition is $s_0 = \phi(s_1, -t_1)$ because

$$\begin{aligned} \phi(s_0, t_1) &= \phi(\phi(s_1, -t_1), t_1) \\ &= \phi(s_1, -t_1 + t_1) = \phi(s_1, 0) = s_1 \end{aligned}$$

where we have used conditions (1) and (2) above. To eliminate the possibility that there is another state that could have been the initial condition suppose that $\phi(s_0', t_1) = s_1$, then

$$\begin{aligned} \phi(s_0', t_1) &= \phi(s_0, t_1) \\ \phi(\phi(s_0', t_1), -t_1) &= \phi(\phi(s_0, t_1), -t_1) \\ \phi(s_0', t_1 - t_1) &= \phi(s_0, t_1 - t_1) \\ \phi(s_0', 0) &= \phi(s_0, 0) \\ s_0' &= s_0 \end{aligned}$$

where we have used conditions (1) and (2) again. Consequently s_0 is the only possible initial condition that leads to state s_1 at time t_1. In other words to find the initial state from the state at time t_1 we treat the state at time t_1 as an initial condition and see where it goes under the dynamical system at time $-t_1$. This type of invertibility is not a general property of dynamical systems but it often does hold and when it does it can be very useful.

Two important and useful ideas in dynamical systems theory are "invariant sets" and "orbits". In the context of invertible dynamical systems we can define an *invariant set* of a dynamical system, $\phi : S \times \mathbf{R} \to S$, to be a non-empty subset $U \subset S$ with the property that for every $s \in U$ and every $t \in \mathbf{R}$ it follows that $\phi(s, t) \in U$. The *orbit* of a point $s_0 \in S$ is the set $\{\phi(s_0, t) : t \in \mathbf{R}\}$.

The whole state space, S, is a trivial example of an invariant set. The orbit of a point is another example of an invariant set. To see this, let $s_1 \in S$ be some point in the orbit of $s_0 \in S$. Then there must be some time $t_1 \in \mathbf{R}$ such that

$\phi(s_0, t_1) = s_1$ and by invertibility $\phi(s_1, -t_1) = s_0$. So for any $t \in \mathbf{R}$ the state

$$\begin{aligned} \phi(s_1, t) &= \phi(s_1, t_1 + (t - t_1)) \\ &= \phi(\phi(s_1, t_1), t - t_1) \\ &= \phi(s_0, t - t_1) \end{aligned}$$

is in the orbit of s_0. The orbit of s_0 is therefore an invariant set. It also follows that s_0 is in the orbit of s_1 and that the orbit of s_1 is the same as the orbit of s_0. The fact that the orbit of a point is an invariant set means that it doesn't matter which point of an orbit we use to specify it.

It should be clear that no proper subset of an orbit can be an invariant set. Orbits can not be decomposed into smaller invariant sets and in that sense orbits are the smallest type of invariant set. The whole state space is the largest invariant set. There are types of invariant sets between these two extremes.

It should also be clear that the union of a collection of invariant sets forms an invariant set so of course the union of a collection of orbits is an invariant set. Since orbits are the smallest type of invariant set it would be nice if we could decompose every invariant set into a collection of orbits. This is indeed the case. Every invariant set is partitioned by the orbits within it. Establishing this fact is especially easy for invertible dynamical systems.

To see this fact we first verify that distinct orbits must be disjoint sets. Suppose the orbits of $s_1, s_2 \in S$ have a point, s_0, in common. Since s_0 is in the orbit of s_1 the orbits of s_0 and s_1 are the same. Since s_0 is in the orbit of s_2 the orbits of s_0 and s_2 are the same. Therefore the orbits of s_1 and s_2 must be the same. If two orbits are not disjoint then they are really the same orbit.

Every point in an invariant set is in its own orbit. The orbit of a point in an invariant set can not go outside of the invariant set since that would contradict the set being invariant. Therefore the collection of the orbits for all of the points in an invariant set forms a partition of the invariant set.

Since the entire state space is an invariant set it too can be partitioned into orbits. The collection of all orbits contains a great deal of information about the dynamical system and is sometimes called a *phase portrait*. For one, two, and three dimensional dynamical systems the visual portrayal of some orbits in the phase portrait of a dynamical system, along with arrows indicating the future direction along an orbit, can be very useful for understanding dynamical systems. Images of phase portraits are used throughout this text.

Any union of orbits is an invariant set and so there are many types of invariant sets most of which are not useful to consider. In almost every case the invariant sets which are chosen for study form a single connected set. In addition to being connected the invariant sets that merit attention usually have other useful topological properties such as being open or closed sets.

The restriction of a dynamical system to an invariant set is itself a dynamical system. The study of complicated dynamical systems which arise in scientific applications often involves recognizing important invariant sets on which the dynamics is not so complicated. We will have more to say about this later. For now let us bring up one of the most important types of invariant sets. A point $s_0 \in S$ is a *fixed point* (or *equilibrium point*) if for all $t \in \mathbf{R}$ it follows that $\phi(s_0, t) = s_0$. The orbit of a fixed point is a single point. One reason why fixed points are important is because it often occurs that the state of the system tends towards a fixed point.

We will also make use of the idea of a parameterized family of dynamical systems. A *parameterized family of dynamical systems* is a continuous map $\phi : S \times \mathbf{R} \times \mathbf{R}^n \to S$ with the property that for each $\mu \in \mathbf{R}^n$ the map $(s, t) \mapsto \phi(s, t, \mu)$ is a dynamical system. μ is called the *parameter(s)* and its value is fixed for any $s \in S$ and all $t \in \mathbf{R}$. Different values for the parameters, μ, typically yield different dynamical systems. One can abstractly consider the effect of changing the value of the parameters but formally the parameters' value do not depend on the time space. There is an informal exception to fixing parameter values which is called the "quasistatic variation of parameters", which is discussed below.

If, when varying the parameter of a parameterized family of dynamical systems, a topologi-

cal change occurs, (*e.g.* a change in the number of fixed points occurs) then the family of dynamical systems is said to have undergone a bifurcation. We will discuss bifurcations in more detail through the upcoming examples.

In many cases dynamical systems arise as the solutions to a system of ordinary differential equations (ODEs) but we should be careful to distinguish between the two types of mathematical objects. Not all differential equations are meant to model dynamical processes. For example a differential equation can be used to give us the shape of a motionless hanging cable. Even when an ODE is derived with the intention of modeling some naturally occurring time dependent process it is possible that the set of solutions will fail to satisfy the conditions of a dynamical system (even using most of the broader definitions of a dynamical system).

There is, however, the well known "existence and uniqueness" theorem for ODEs. This theorem provides a fairly simple test which is almost enough[8] to ensure that the set of solutions to a particular system of ordinary differential equations will form a dynamical system as we have defined it here. For simplicity our examples will come from solutions to ODEs which do satisfy our simple definition of a dynamical system.

ODEs are useful for modeling dynamical processes mainly because they have solutions that form dynamical systems. It is the dynamical systems that are of interest to us. It is generally straightforward to check that a particular dynamical system is a solution for a given ODE. On the other hand given an ODE which satisfies the existence and uniqueness theorem it is often very difficult to find anything more than an approximate solution. Even centuries old problems such as the motion of celestial bodies are the subject of continuing research. In this text, we will not consider methods for finding solutions or approximate solutions to ODEs in detail. Our goal is to elucidate a framework for the study of embodied cognition. Dynamical systems theory forms the basis for that framework.

[8]The existence and uniqueness theorem does not guarantee that the solutions of an ODE which satisfies its test will be defined for all $t \in \mathbf{R}$ but only for all t in some open interval of $\mathbf{R}$ which contains 0. Rather than use a more complicated definition for dynamical systems we will avoid examples where this limitation occurs.

2.2 Dynamical systems open to an environment

One simple way to represent embodied cognition mathematically is by having a state space for an agent and a state space for the environment the agent is embedded in. The agent state space will be denoted by S_α and the environmental state space will be denoted by S_ϵ. The Cartesian product $S_\epsilon \times S_\alpha$ will be called the *total state space* and it will be denoted by S_τ. We will write the members of S_τ as ordered pairs $(\mathbf{r}, \mathbf{x})$ where $\mathbf{r} \in S_\epsilon$ and $\mathbf{x} \in S_\alpha$.

The environment and the agent influence each other's behavior and we represent their combined behavior with a dynamical system on S_τ. We denote this dynamical system by $\phi_\tau : S_\tau \times \mathbf{R} \to S_\tau$.

We are also interested in what goes on "inside" the agent in a changing environment. We are particularly interested in the how the internal state of the agent responds to changing sensory input from the environment (the environment is also influenced by the agent, but we will not focus on that direction of coupling in this text). However, we also want to consider what happens when we put an agent into different environments, put different agents in the same environment, or otherwise permute the associations between agents and environments. Finally, we want to be able to compare the intrinsic dynamics of an agent with the impact of an environment upon that agent. Consequently we introduce machinery to describe the behavior of the agent both when it is in an environment and when it is decoupled from that environment.

We represent the intrinsic dynamics of the agent with a dynamical system on S_α. We denote this dynamical system by $\phi_\alpha : S_\alpha \times \mathbf{R} \to S_\alpha$.

We introduce a third dynamical system $\varphi_\tau : S_\tau \times \mathbf{R} \to S_\tau$ whose purpose is to specify how the dynamical systems ϕ_τ and ϕ_α are related. Roughly speaking the effect of φ_τ on the total state space is to leave the environmental component unchanged and to change the agent compo-

nent as though it were still isolated. Metaphorically speaking we can think of this as severing any coupling between the agent and the environment and placing the agent in an environment which has been frozen in time. More precisely stated, for each $\mathbf{r} \in S_\epsilon$, $\mathbf{x} \in S_\alpha$, and $t \in \mathbf{R}$ we let

$$\varphi_\tau((\mathbf{r}, \mathbf{x}), t) = (\mathbf{r}, \phi_\alpha(\mathbf{x}, t))$$

We can easily check that this definition for the map φ_τ satisfies the properties for being a dynamical system. Essentially φ_τ inherits the properties of a dynamical system from ϕ_α. For condition (1)

$$\varphi_\tau((\mathbf{r}, \mathbf{x}), 0) = (\mathbf{r}, \phi_\alpha(\mathbf{x}, 0)) = (\mathbf{r}, \mathbf{x})$$

And for condition (2)

$$\begin{aligned} \varphi_\tau((\mathbf{r}, \mathbf{x}), t_1 + t_2) &= (\mathbf{r}, \phi_\alpha(\mathbf{x}, t_1 + t_2)) \\ &= (\mathbf{r}, \phi_\alpha(\phi_\alpha(\mathbf{x}, t_1), t_2)) \\ &= \varphi_\tau((\mathbf{r}, \phi_\alpha(\mathbf{x}, t_1)), t_2) \\ &= \varphi_\tau(\varphi_\tau((\mathbf{r}, \mathbf{x}), t_1), t_2) \end{aligned}$$

There is in dynamical systems theory a more concise way to define φ_τ using the idea of a product dynamical system. For a pair of dynamical systems

$$\phi_1 : S_1 \times \mathbf{R} \to S_1$$
$$\phi_2 : S_2 \times \mathbf{R} \to S_2$$

we define the *product dynamical system*

$$\phi_1 \times \phi_2 : S_1 \times S_2 \times \mathbf{R} \to S_1 \times S_2$$

to be

$$(\phi_1 \times \phi_2)(s_1, s_2, t) = (\phi_1(s_1, t), \phi_2(s_2, t))$$

for any $s_1 \in S_1$, $s_2 \in S_2$, and $t \in \mathbf{R}$. The reader can check that the map $\phi_1 \times \phi_2$ satisfies the two conditions for being a dynamical system.

To use this definition for a product dynamical system we let $\iota_\epsilon : S_\epsilon \times \mathbf{R} \to S_\epsilon$ denote the identity dynamical system on S_ϵ (*i.e.* the dynamical system which keeps every point in S_ϵ fixed for all time). It then follows that $\varphi_\tau = \iota_\epsilon \times \phi_\alpha$.

Geometrically the space S_τ is partitioned into an infinite number of "parallel" copies of S_α and on each copy of S_α the dynamics under φ_τ is the same as on S_α (see Figure 1). The dynamical system φ_τ leaves each subset of the form $\{\mathbf{r}\} \times S_\alpha$ invariant where $\mathbf{r} \in S_\epsilon$. And the action of φ_τ on each subset $\{\mathbf{r}\} \times S_\alpha$ is essentially the same as the action of ϕ_α on S_α.

We now consider the coupling between the environment and the agent. This is represented mathematically by specifying a continuous deformation of φ_τ into ϕ_τ. More precisely stated we specify a continuous function of the form

$$\Phi : S_\tau \times \mathbf{R} \times [0, 1] \to S_\tau$$

which satisfies

$$\begin{aligned} \Phi(\mathbf{r}, \mathbf{x}, t, 0) &= \varphi_\tau(\mathbf{r}, \mathbf{x}, t) \\ \Phi(\mathbf{r}, \mathbf{x}, t, 1) &= \phi_\tau(\mathbf{r}, \mathbf{x}, t) \end{aligned}$$

The map Φ is an example of a homotopy. In this context the number $\mu \in [0, 1]$ provides a measure for the degree of coupling of the environment with the agent. The number 0 corresponds to no coupling and the number 1 corresponds to full coupling.

For a fixed $\mu \in (0, 1)$ we do not require $\Phi(\mathbf{r}, \mathbf{x}, t, \mu)$ to be a dynamical system although in actual practice it usually is. We are more interested in comparing and contrasting the internal dynamics of an isolated agent, ϕ_α, with the dynamics of an agent in an environment, ϕ_τ. As noted above, the map φ_τ can be thought of as an idealized situation in which the agent has been decoupled from the environment and placed into an environment which has been frozen in time. Increasing μ above 0 can be thought of as coupling the agent to the environment and unfreezing time but we don't want to imply that there is any real physical meaning to this act. In particular, we will not require the intermediate stages of the homotopy (*i.e.* $\mu \in (0, 1)$) to represent actual states of coupling that occur between the environment and the agent. We only require that the act of coupling result in a continuous deformation from the dynamical system $\iota_\epsilon \times \phi_\alpha$ to the dynamical system ϕ_τ. The map Φ specifies a particular continuous deformation which represents a particular way of coupling the agent to an environment. It is worth noting however

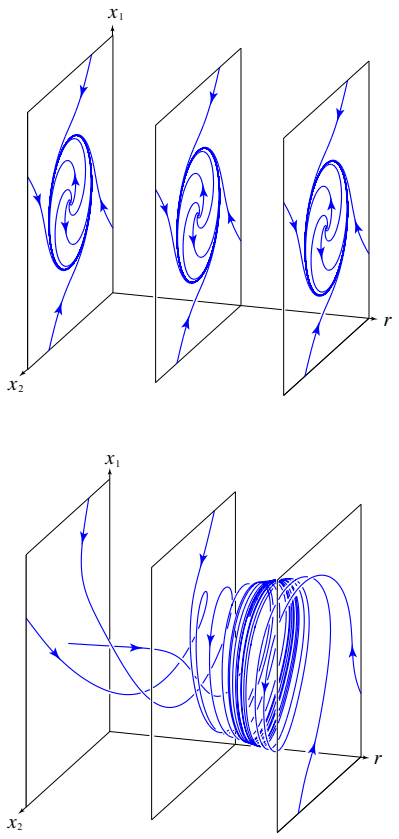

Figure 1: The top panel shows a phase portrait for a dynamical system of the form $\varphi_\tau = \iota_\epsilon \times \phi_\alpha$. In this example the state space for the environment is one dimensional. The dynamical system ϕ_α for the agent is two dimensional and has an attracting periodic orbit which is circular in shape. The bottom panel shows the phase portrait for a dynamical system, ϕ_τ, which has been obtained by a continuous deformation of φ_τ. The diagram as a whole illustrates the basic elements involved in modeling embodied cognition: a total system ϕ_τ (bottom panel) which models the behavior of an agent and environment coupled together, an agent system ϕ_α (frames of top panel) which shows how the agent behaves when it is decoupled from the environment, and a modified total system φ_τ (top panel) which shows how the agent system and total system are related.

that in those cases where $\Phi(\mathbf{r}, \mathbf{x}, t, \mu)$ is in fact a dynamical system for all $\mu \in [0, 1]$ that we can think of $\Phi(\mathbf{r}, \mathbf{x}, t, \mu)$ as a one parameter family of dynamical systems with parameter μ.

Generally none of the subsets $\{\mathbf{r}\} \times S_\alpha \subset S_\tau$ will be invariant under ϕ_τ. None of the orbits of the total dynamical system need to be confined within some copy of the agent's state space. However since the total state is a pair of states for the environment and the agent no matter what state the total system is in the agent must be in some state of its own. We denote the projection $(\mathbf{r}, \mathbf{x}) \mapsto \mathbf{x}$ from the total state space to the agent's state space by π. It is often useful in dynamical systems theory to project orbits in a variety of ways, to facilitate analysis. However, we distinguish the projection π because the images of orbits (which we will call "paths") are the states that actually occur in an agent embedded in an environment.

The preimage of an agent's state, $\mathbf{x} \in S_\alpha$, under π is the set of all states of the total system that give rise to the state $\mathbf{x}$ in the agent. The map π is generally not one to one and there will usually be many states of the total system that produce the same state in the agent.

When the agent is open to its environment we run the dynamical system ϕ_τ and project the state of the total system to the state of the agent. The projection of the dynamics in S_τ does not in general lead to a dynamical system on S_α. For example, two orbits of ϕ_τ can project to two curves which cross each other. An agent whose current state is at a crossing point has more than one possible future which depends on the state of the environment at that moment.

We summarize these assumptions into a formal system which we call an *open dynamical system.* For the purposes of this text, an open dynamical system consists of a pair of state spaces S_ϵ, S_α, a time space $\mathbf{R}$, the interval $[0, 1]$, and a continuous map

$$\Phi : S_\epsilon \times S_\alpha \times \mathbf{R} \times [0, 1] \to S_\epsilon \times S_\alpha$$

which satisfies the following conditions. For all $\mathbf{r} \in S_\epsilon$, $\mathbf{x} \in S_\alpha$, and $t \in \mathbf{R}$ the map

$$(\mathbf{r}, \mathbf{x}, t) \mapsto \Phi(\mathbf{r}, \mathbf{x}, t, 0)$$

is a dynamical system of the form

$$\iota_\epsilon \times \phi_\alpha : S_\epsilon \times S_\alpha \times \mathbf{R} \to S_\epsilon \times S_\alpha$$

and the map

$$(\mathbf{r}, \mathbf{x}, t) \mapsto \Phi(\mathbf{r}, \mathbf{x}, t, 1)$$

is a dynamical system of the form

$$\phi_\tau : S_\epsilon \times S_\alpha \times \mathbf{R} \to S_\epsilon \times S_\alpha$$

This concept of an open dynamical system is not the only formal system that can be used to treat embodied cognition but it is one of the simplest (in separate work we consider a more general formulation for an open dynamical system).

Because more assumptions go into the definition of an open dynamical system versus just a dynamical system it might seem that the class of open dynamical systems is smaller than the class of dynamical systems, in the same way that, for example, mammals are a more restricted group of organisms than animals are because there are more requirements for an organism to be a mammal. However, formally dynamical systems can be thought of as special cases of open dynamical systems.

We can see this as follows. Abstractly speaking we can regard any dynamical system as a model for some type of agent. Given a state space, S_α, and a dynamical system, ϕ_α, we can let S_ϵ be the single point set $\{0\}$. There is next to no environment to speak of and no way for the environment to change state. The state space $S_\tau = \{0\} \times S_\alpha$ is just a copy of S_α. The map $\pi : S_\tau \to S_\alpha$ is a one to one correspondence. Let $\Phi((0, \mathbf{x}), t, \mu) = (0, \phi_\alpha(\mathbf{x}, t))$ for all $\mu \in [0, 1]$. The action of ϕ_τ on S_τ is essentially the same as the action of ϕ_α on S_α. This open dynamical system is virtually nothing more than just a dynamical system. We are just prefixing agent states with a '0' to get total states. Every dynamical system can be used to define an open dynamical system in this way. This procedure for turning a dynamical system into an open dynamical system never turns two distinct dynamical systems into the same open dynamical system. Every open dynamical system with $S_\tau = \{0\} \times S_\alpha$ and $\Phi((0, \mathbf{x}), t, \mu) = (0, \phi_\alpha(\mathbf{x}, t))$ for all $\mu \in [0, 1]$ can be thought of as being just a dynamical system $\phi_\alpha : S_\alpha \times \mathbf{R} \to S_\alpha$. This allows us to regard dynamical systems as a special case of open dynamical systems. We can think of the act of enlarging the total space so that the projection is no longer one to one as the act of opening the dynamical system up to some environment.[9]

In what follows, when we refer to "open dynamical systems" we typically mean open dynamical systems for which the projection map is not the identity map (i.e. which are not essentially just copies of dynamical systems), what might be thought of as "proper open dynamical systems."

2.3 Paths and Isotypes

We call the image of an orbit of a total system ϕ_τ to an agent space S_α under π a *path*. Paths are important objects, from the standpoint of cognitive science, because they correspond to what happens inside an agent when it is embedded in an environment. Ultimately, we think it will be possible to analyze representational processes in embodied agents by studying the topology and geometry of these projected orbits. Hence, it is important to cognitive research that it have tools for analyzing them. In this section we introduce some of these tools.

The issue is complicated by the fact that paths are not as simple to analyze as orbits, because they can cross. The orbits of a dynamical system form a partition of the state space. This fortunate situation helps make dynamical systems very amenable to analysis. Matters are not as simple with open dynamical systems. In an open dynamical system we run the dynamical system

[9] Let us now consider open loop control systems again. Control systems are parameterized systems of differential equations $\dot{\mathbf{x}} = f(\mathbf{x}, \mathbf{u})$ where $\mathbf{x}$ is usually a point on a manifold and $\mathbf{u}$ is usually a point in a convex set. The open loop problem is to find a way to vary $\mathbf{u}$ over time so that $\mathbf{x}$ will go from point A to point B. This is similar to an open dynamical system in that we can think of $\mathbf{x}$ as the agent's state and the pair $(\mathbf{x}, \mathbf{u})$ the total state. However in open loop problems there is no dynamical system on the total state space. We are free to vary $\mathbf{u}$ as we like and we try to extend this ability to vary $\mathbf{x}$ as we like.

ϕ_τ on the total state space S_τ and project down to the agent state space S_α. The projected image of an orbit can cross itself as well as the projected image of other orbits, so the description of what is happening inside of S_α is more complicated. Thus, the state space S_α is completely covered by paths but the collection of all paths in S_α does not in general form a partition of S_α.

We begin by contrasting the generic behaviors that will actually be observed in an open dynamical system, with exceptional behaviors that are not normally observed. We can do this by formalizing a method of analysis which is commonly employed in classic dynamical systems theory in an informal manner. It is very common for neighboring orbits of a dynamical system to be homeomorphic, that is to have the same topology. When this occurs the behavior of a dynamical system is qualitatively the same when initial conditions change slightly. These are the orbits one will normally observe, since in actual practice, even if initial conditions are not known with perfect precision, neighboring orbits are qualitatively the same. On the other hand, there are orbits which are topologically different from every neighboring orbit and either the exceptional character of such orbits aids in the understanding of the dynamical system or else such orbits tend to not be relevant in practice. In the agent space it is also true that neighboring paths are often homeomorphic to each other though they can also be exceptional. To understand the behavior of an open dynamical system it is useful to identify these collections of homeomorphic paths (which we call "generic paths") as well as the exceptional paths.

We first define what we mean by a generic path. Generic paths are the paths that will actually be observed in an agent, since they correspond to dense collections of orbits in the total space all of which produce topologically the same kind of path. To check if a path p is generic, we consider its preimage under π. If there exists some orbit q in this preimage, and an open set $U \subset S_\tau$ which contains q, and every orbit in U projects to a path which is topologically the same as p, then p is a *generic* path. If a path is not generic then it is *exceptional*.

Generic paths can be grouped into collections of paths which do roughly the same thing. 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 such that every intermediate stage is a generic path then we say the paths p_1, p_2 are *isotypic* or that they are members of the same *isotype*. Being isotypic is an equivalence relation on the collection of generic paths in S_α, so that the isotypes of a system provides an essentially complete classification of the typical behaviors of that system. The union of all of the points in an isotype is an *isotype region*.

The isotypic and exceptional paths of a system can be classified according to their topology, which, as we will see, can be interesting from the standpoint of the representational processes that occur in agents.

In dynamical systems theory attractors are often regarded as the behavior that a dynamical system tends to exhibit. This is because over time the state of the system ends up inside a small neighborhood of the attractor. Strictly speaking, attractors are exceptional; unless the system starts out in the attractor it never actually reaches it. For example the fixed points of most dynamical systems which arise in applications are isolated and qualitatively different from all of their nearby orbits. However, attractors can still tell us about the normal behavior of the system even if they do not themselves correspond to normal behavior. Thus, it is important to look both at attracting sets, and at orbits which tend towards attracting sets (henceforth "basin orbits"). The set of all basin orbits of an attracting set is its *basin of attraction*. In models of cognitive agents we will clearly want to see what the normal behaviors are. Thus, it will help to have open analogue both for orbits in attractors and for basin orbits.

Attractors are a particular type of attracting set. We define an attracting set as follows. A subset $A \subset S$ of the state space of a dynamical system $\phi : S \times \mathbf{R} \to S$ is an *attracting set* if it is closed, invariant, and there is an open set U containing A with the property that for all $s \in U$ the distance between $\phi(s,t)$ and A goes to zero as t goes to positive infinity. An *attractor* is an

attracting set which contains a dense orbit. This gives an attractor a kind of unity that an attracting set need not have. Any orbit which enters U is a basin orbit, and can be thought of as being "trapped" in U. Similarly a subset $A \subset S$ is a *repelling set* if it is closed, invariant, and there is an open set U containing A with the property that for all $s \in U$ the distance between $\phi(s,t)$ and A goes to zero as t goes to negative infinity. A *repeller* is a repelling set which contains a dense orbit.

The analogue for open dynamical systems of an orbit of an attracting set is an *attracting path*. The projection of an orbit in an attracting set A under π is an attracting path. The open analogue of a basin orbit is a *trapped path*, and it is simply the projection of a basin orbit in the total space.

Aside from some very exceptional cases, all of the paths of an isotype will either be trapped or not trapped (because of the requirement that two paths of an isotype must be isotopic through generic paths). When all of the paths of an isotype are trapped we say that the *isotype is trapped* or that we have a *trapped isotype*. The typical behavior in the agent's state space of an open dynamical system whose total dynamical system has an attracting set is characterized by the trapped isotypes.

2.4 Introductory examples

In this section we show how the concepts defined above work in simple, familiar cases, and thereby begin to show how they can be put to use in cognitive research. We consider three cases. First, the classic pendulum, which shows how isotypes can be used to classify the basic behaviors of a system. Second, the damped pendulum, which illustrates the concepts of attracting paths, trapped paths, and trapped isotypes. Third, we consider the classical example of a hysteresis loop, contrasting a traditional analysis with an analysis in terms of open systems. This demonstrates some of the advantages of the concept of an open system. We assign colors to paths that correspond to particular isotypes or classes of attracting path.

For simplicity the first two examples, the undamped and damped pendulum, will not be proper open dynamical systems. Recall that this means that the total space, S_τ, is the same as the agent state space, S_α and the projection π is the identity map. The projection of an orbit to a path does not change it. The paths of this open dynamical system are in fact the orbits of the dynamical system. The classification of paths into isotypes is therefore also a classification of orbits of the dynamical system.

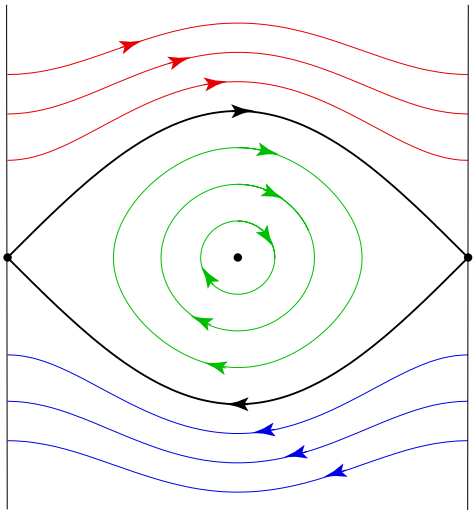

Figure 2: Phase portrait of the undamped pendulum. The cylindrical state space is obtained by identifying left and right edges. The horizontal direction represents the angular position of the pendulum and the vertical direction represents the angular velocity. The saddle point is shown on both edges. The black orbits are exceptional. The green orbits are part of the isotype for the pendulum to swing. The red orbits are part of the isotype for the pendulum to rotate clockwise. The blue orbits are part of the isotype for the pendulum to rotate counter-clockwise.

In a sense the undamped pendulum is a classic example of the classification of the orbits of a dynamical system into isotypes. Although this classification has not been performed with the formalism of open dynamical systems the isotypes are easily recognized and they have been pointed out many times.

The bob of the pendulum moves along a cir-

cle and the pendulum's state is given by its angular position and its angular velocity. The pendulum's state space is the Cartesian product of $\mathbf{S}^1 \times \mathbf{R}$ *i.e.* a cylinder (see Figure 2). The dynamical system has exactly two fixed points in the cylinder. One fixed point corresponds to the pendulum hanging straight down. In a neighborhood around this fixed point it is surrounded by concentric circularly shaped orbits. This type of fixed point is usually called a "center" in dynamical systems theory. The other fixed point corresponds to the pendulum standing straight up. This fixed point is the boundary point of a pair of orbits that correspond to the two directions that a pendulum can fall from a nearly upright position and rise back up to a nearly upright position. This type of fixed point is usually called a "saddle point". Roughly speaking, we say that orbits which begin and end at the same fixed point are "homoclinic". Every other orbit near the saddle point simply pass by it.

The two fixed points and the two homoclinic orbits are the four exceptional paths. The complement of the four exceptional paths consists of three connected components. The connected components are covered by generic paths. Once an undamped pendulum is set into motion the path that it follows will almost always be generic. As can be seen from the phase portrait in Figure 2 the paths in each connected component form an isotype and the connected components are isotype regions. One isotype region corresponds to the pendulum swinging back and forth without ever completing a revolution. A second isotype region corresponds to the pendulum rotating in a clockwise direction. The third isotype region corresponds to the pendulum rotating in a counter-clockwise direction.

This analysis of the undamped pendulum is well known. Similar analyses have been performed with other dynamical systems[10]. This type of classification is often done heuristically on low dimensional dynamical systems without the use of a formal procedure.

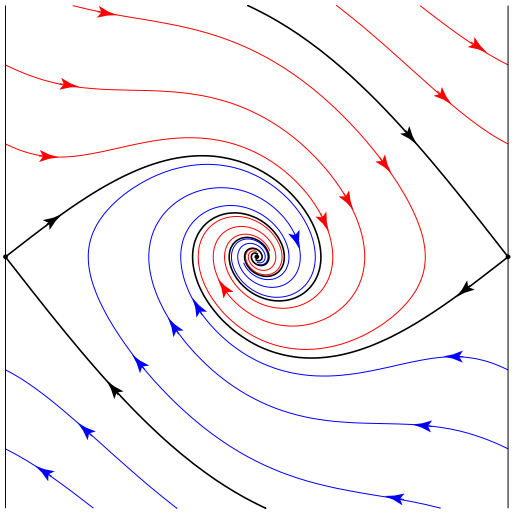

Figure 3: Phase portrait of the damped pendulum. The cylindrical state space is obtained by identifying left and right edges. The saddle point is shown on both edges. The black orbits are exceptional. The red orbits are part of the isotype for the pendulum to rotate clockwise for a while and then swing back and forth with decaying amplitude. The blue orbits are part of the isotype for the pendulum to rotate counter-clockwise for a while and then swing back and forth with decaying amplitude.

The classification of the orbits of a dynamical system into isotypes has two advantages over the classification of the paths of a proper open dynamical system. The topology of orbits tends to be simpler than that of paths and orbits form a partition of the state space while paths generally do not.

Next we consider what happens with the paths in the pendulum's state space when its motion is damped by friction, which illustrates the concepts of an attracting path, a trapped path, and a trapped isotype. The dynamics of the damped pendulum differ from the undamped pendulum in that a damped pendulum has an attractor while the undamped pendulum does not (see Figure 3). The analysis of the damped pendulum also has been performed many times and is not

[10] For example this classification of orbits for the pendulum is similar to a method known as a cell-decomposition by separatrices. This type of decomposition is generally performed with two dimensional dynamical systems to ascertain whether they have a property known as structural stability (DeBaggis [1952]; Peixoto [1959]; Hirsch & Smale [1974]).

difficult. The state space is the same. There are six exceptional paths. Two exceptional paths are fixed points which are in the same location as in the undamped case and have the same meaning for the pendulum.

One fixed point corresponds to the hanging down position. In a neighborhood of this fixed point the state of the system spirals in towards it. This type of fixed point is often called a "stable focus" in dynamical systems theory. It is an attracting path and the only attracting path for this system.

The other fixed point corresponds to the standing upright position for the pendulum. This point is still a saddle point. Each of the two homoclinic orbits of the undamped pendulum "slivers apart" to give four orbits for the damped pendulum. These four orbits are the remaining four exceptional paths. Two of these paths correspond to the pendulum going from a nearly upright position to a nearly downward position. Roughly speaking orbits which begin at one fixed point and end at another fixed point are usually called "heteroclinic orbits" in dynamical systems theory. The remaining two exceptional paths of the damped pendulum correspond to the pendulum rotating and coming to nearly rest in an upright position.

This system has two trapped isotypes, which are the connected components formed by the complement of the exceptional paths. These are the two types of behavior that will typically be observed in this system. They correspond to the two directions of rotation of the pendulum, from which it will slow down and eventually oscillate back and forth with a decaying amplitude.

The addition of friction to the undamped pendulum results in the oscillating isotype of the undamped pendulum splitting into two parts with one part merging with the rotate clockwise isotype of the undamped pendulum and the other part merging with the rotate counter-clockwise isotype of the undamped pendulum. Both isotypes of the damped pendulum are trapped by the attracting path.

We now consider an example which is a proper open dynamical system. Systems with overlapping paths have been analyzed many times in a variety of contexts and good examples of overlapping paths can be found in the study of hysteresis loops. Hysteresis loops arise in a wide variety of circumstances where there are fast and slow variables (including various contexts in cognitive science, *e.g.* Schmidt *et al.* [1990]; Tuller *et al.* [1994]). Hysteresis loops are often analyzed using a quasistatic variation of parameters approach. In this method one or more slow variables are treated as fixed parameters, the subsystem of fast variables is analyzed, and then the parameters are allowed to vary again. This approach can be useful but it runs in to a number of technical difficulties as we show. One solution to these difficulties is the formalism of open dynamical system.

To show this, we construct a fairly simple example exhibiting hysteresis and contrast an analysis of the system using quasistatic variation of parameters with an analysis of the same system considered as an open dynamical system. The analysis of the system as an open dynamical system also shows, in an idealized way, how open dynamical systems can be used to model embodied cognition.

The state space S_α will be $\mathbf{R}^2$. The dynamical system for the agent, ϕ_α, will be given by the solutions to the ODE

$$\begin{aligned} \dot{x_1} &= 0 \\ \dot{x_2} &= 2x_1 - x_2(x_2^2 - 3). \end{aligned} \tag{3}$$

Its phase portrait is shown in Figure 4. The variable x_1 can be thought of as a sensor of the agent and x_2 can be thought of as the agent's processing of the sensor's state. We can further decompose ϕ_α into a set of parallel one dimensional systems each corresponding to the sensor value, x_1, clamped at different levels. The system ϕ_α will settle into a fixed point attractor.

If x_2 is increasing then $\dot{x_2} > 0$, if x_2 is decreasing then $\dot{x_2} < 0$ and if x_2 is at a fixed point of a one dimensional subsystem then $\dot{x_2} = 0$.

The set of all ordered pairs (x_1, x_2) that satisfy the equation $\dot{x_2} = 0 = 2x_1 - x_2(x_2^2 - 3)$ forms a curve. For each value of x_1 this curve gives the locations of the fixed points for the corresponding one dimensional subsystem. This curve is

shown in Figure 4. Note that since the value of x_1 is essentially clamped in our agent system whenever x_2 is at a fixed point of the one dimensional subsystem the pair (x_1, x_2) is at a fixed point of ϕ_α. Thus the curve given by the equation $0 = 2x_1 - x_2(x_2^2 - 3)$ is the set of fixed points for the agent dynamical system.

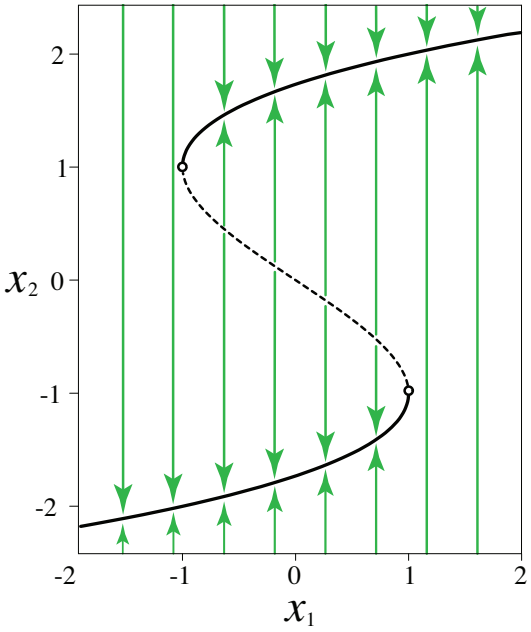

Figure 4: A phase portrait for the agent dynamical system in equation (3). The locus of fixed points of ϕ_α is the black curve. The vertical green lines illustrate some of the other orbits of ϕ_α.

For the one dimensional subsystems bifurcations occur when there is a change in the number of its fixed points as x_1 varies. For $x_1 < -1$ there is only one fixed point and it is located in the negative real numbers. For $x_1 < -1$ and x_2 above the fixed point the value of $2x_1 - x_2(x_2^2 - 3)$ is negative and x_2 decreases with time. For $x_1 < -1$ and x_2 below the fixed point $2x_1 - x_2(x_2^2 - 3) > 0$ and x_2 increases.

For $x_1 = -1$ a new fixed point appears at $x_2 = 1$. For x_1 slightly greater than -1 there are two fixed points in the positive real numbers and one fixed point in the negative real numbers. The reader can check that the value of x_2 moves away from the middle fixed point and towards one of the outer fixed points by computing the sign of $2x_1 - x_2(x_2^2 - 3)$ in between the fixed points.

As x_1 is increased the two positively located fixed points spread apart and when $x_1 = 0$ one fixed point lies on the origin, one fixed point is in the positive reals, and one fixed point is in the negative reals. After x_1 passes through zero one fixed point is located in the positive reals and two fixed points are located in the negative reals. When $x_1 = 1$ the two fixed points in the negative reals merge into a single fixed point while a second fixed continues to exist in the positive reals. After x_1 passes through 1 there is only a single fixed point and it is located in the positive real numbers.

The type of bifurcation this parameterized family of dynamical systems undergoes at both $x_1 = \pm 1$ is known as a saddle node bifurcation. Saddle node bifurcations are characterized by either the appearance or disappearance of a pair of fixed points.

The environmental state space S_ϵ will be $\mathbf{R}$. The total state space is then $\mathbf{R}^3$ and the map π is an orthogonal projection of $\mathbf{R}^3$ to $\mathbf{R}^2$. The total dynamical systems ϕ_τ will be the solutions to

$$\begin{aligned} \dot{r} &= -\omega x_1 \\ \dot{x_1} &= \omega r \\ \dot{x_2} &= 2x_1 - x_2(x_2^2 - 3) \end{aligned} \tag{4}$$

with $\omega > 0$. The variable r stands for the state of an environment. The variables r, x_1 do not depend on x_2. The differential equations for r, x_1 force them to vary sinusoidally. We can think of this as the environment and sensor interacting to set up an oscillation between themselves. The differential equation for x_2 depends explicitly on x_1 but not on r. The agent responds to the oscillations that appear when it is placed in the environment by processing just information that it gets from the sensor. For small values of ω the variables r, x_1 change slowly in comparison to x_2.

We can let the dynamical system φ_τ be the solutions to ODE (4) with $\omega = 0$. The dynamical system ϕ_τ can be continuously deformed to φ_τ simply by letting ω go to zero.

We perform a quasistatic variation of parameter analysis to approximate the changes of state in S_α produced by ϕ_τ. When ω is small enough we can approximate the time course of x_2 by treating x_1 as a fixed parameter. By fixing x_1 we are back to the single variable ODE $\dot{x_2} = 2x_1 - x_2(x_2^2 - 3)$ which we have already analyzed.

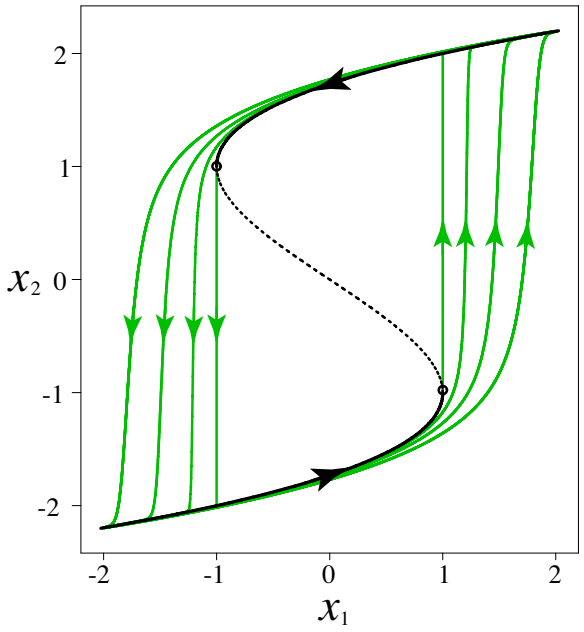

Figure 5: The hysteresis loops for system (4) are shown in green. The set of fixed points for ϕ_α is shown by the black curve. The innermost hysteresis loop is obtained by quasistatic variation. Continuing outward the hysteresis loops for system (4) are numerically approximated for $\omega = 0.04, 0.16, 0.4$.

The intrinsic dynamics of the agent helps us to understand what happens when the system is embedded in an environment which generates oscillating inputs. When this happens the orbits of the decoupled system become loops (see Figure 5) which in a rough sense follow the set of fixed points of the agent system. It instructive to compare Figures 4 and 5 which contrast the agent's decoupled behavior with its behavior when it is embedded in an environment.

We now let x_1 slowly vary over time. For simplicity the way we will vary x_1 is by having it oscillate about the value 0 with an amplitude of 2. And we suppose that it starts out with the value -2. With this initial condition and with ω sufficiently small x_2 has time to become very close to the unique attracting fixed point. As x_1 slowly increases the location of the attracting fixed point slowly increases and the value of x_2 closely follows the attracting fixed point. After x_1 passes through -1 a second attracting fixed point appears through a saddle node bifurcation but x_2 remains in the basin of attraction for the lower attracting fixed point and so it continues to follow the lower attracting fixed point.

However when x_1 passes through 1 the lower fixed point disappears through another saddle node bifurcation and x_2 proceeds towards the upper attracting fixed point. We suppose that ω is so small that x_2 travels virtually all the way to the upper fixed point before there has been any noticeable change in x_1. On the time scale of x_1 the value of x_2 essentially jumps to the upper fixed point. As x_1 continues to increase x_2 continues to follow the upper attracting fixed point.

When x_1 reaches 2 it turns around and begins to decrease. The value of x_2 keeps on following the upper attracting fixed point because it is the only attracting fixed point around. After x_1 reaches 1 the lower attracting fixed point reappears but x_2 remains in the basin of attraction for the upper attracting fixed point and so it continues to follow the upper attracting fixed point. When x_1 reaches -1 the upper fixed point disappears and the value of x_2 essentially jumps down to the lower attracting fixed point.

In the period of one oscillation in the value of x_1 the points (x_1, x_2) trace out a curve in the plane which is an example of a hysteresis loop. In Figure 5 the attracting fixed points of x_2 are indicated with solid black curves and the jumps from one attracting fixed point to another are shown with vertical green lines. The union of these two solid black curves and two vertical green lines form the kind of ideal version of a hysteresis loop which is often obtained by the quasistatic variation of parameters. These loops are meant to describe the path followed by a system exhibiting hysteresis but such ideal curves are only approximations of what actually happens.

This traditional analysis of hysteresis using

quasistatic variation of parameters has weaknesses in rigor. First, the language employed is kept a little vague in describing what happens: fixed points are said to "move," "appear," and "disappear," even though they are fixed points. Hysteresis loops are described somewhat obscurely as sets of fixed points together with jumps that amount to a kind of teleportation.

Second, the ideal hysteresis loops can only provide an approximate characterization of the loop's shape and its topology. It can be seen in Figure 5 that actual hysteresis loops curl as they approach the branches for the attracting fixed points. Moreover, though it is hard to see in the figures, these paths do not precisely retrace themselves at their ends, and in some cases they can have double points.

Third, types of behavior not captured by the traditional analysis can be observed in systems exhibiting hysteresis. An example of this can be seen in Figure 6. Here we have fixed ω in system (4) at the small value of $1/5$ and we have varied the initial values of r, x_1 in (4) so that the amplitude of oscillation for x_1 decreases from 2 to almost 1. With this frequency and amplitude for x_1 the value of x_2 doesn't have a chance to "jump" to the other attracting fixed point before the missing fixed point "reappears". This type of phenomenon has been observed in other systems (see Schecter [1985]; Wiggins [1990]) but is not often discussed.

To address these issues we can re-examine (4) using the formalism of open dynamical systems. In particular the total dynamical system possesses a useful collection of invariant sets which are the circular cylinders coaxial with the x_2-axis. This is not obvious from just looking at the state space S_α. The square of the distance of a point from the x_2-axis is $r^2 + x_1^2$ and from equation (4) we see that its rate of change is $2r\dot{r} + 2x_1\dot{x_1} = 2r(-\omega x_1) + 2x_1(\omega r) = 0$. So the distance to the x_2-axis does not change. The state of the system always stays inside a cylinder about the x_2-axis.

Furthermore the system's state is always winding around whichever invariant cylinder it is in. This is easiest to see if we consider the projection of the state of the system to the (r, x_1)-plane. If we compute the magnitude of the velocity in the (r, x_1) plane we get $\sqrt{\dot{r}^2 + \dot{x_1}^2} = \sqrt{(-\omega x_1)^2 + (\omega r)^2} = \omega\sqrt{r^2 + x_1^2}$. The speed is ω times the distance of the state from the x_2-axis and so it must be constant. Also the projected state always moves in the direction which is orthogonal to the direction towards the origin. This can be seen by computing the dot product $(\dot{r}, \dot{x_1}) \bullet (r, x_1) = (-\omega x_1, \omega r) \bullet (r, x_1) = 0$.

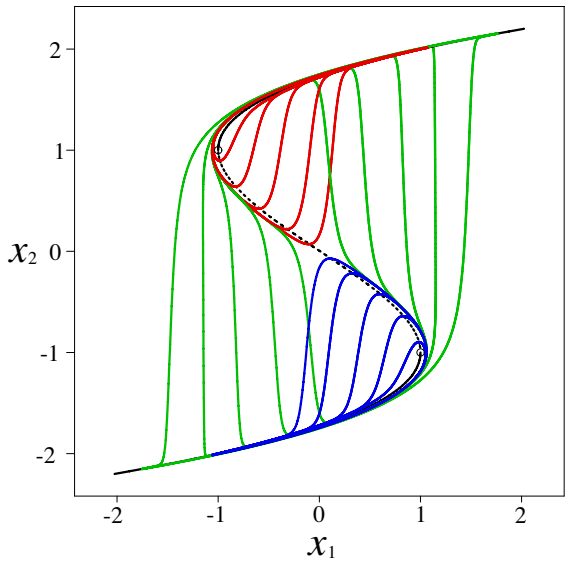

Figure 6: Three types of attracting paths which occur when $\omega = 1/5$. The hysteresis loops are the attracting paths shown in green. When the amplitude of oscillation for x_1 is decreased sufficiently the single symmetrical attracting path becomes two symmetrically positioned attracting paths (shown in red and blue). The curve $\mathbf{S}_\alpha$ where $\dot{x_2} = 0$ in is shown in black.

The projection to the (r, x_1)-plane should not be confused with the projection of the open dynamical system $\pi : S_\tau \to S_\alpha$ which in this case is the projection to the (x_1, x_2)-plane (see Figure 7). We only project to the (r, x_1)-plane in this example as a means to analyze the total dynamical system.

Since every orbit is winding around an invariant cylinder the behavior of this dynamical system is easy to characterize. Orbits can not cross themselves so if an orbit is not periodic then it must progress along the cylinder with each turn. If the future states of a non-periodic

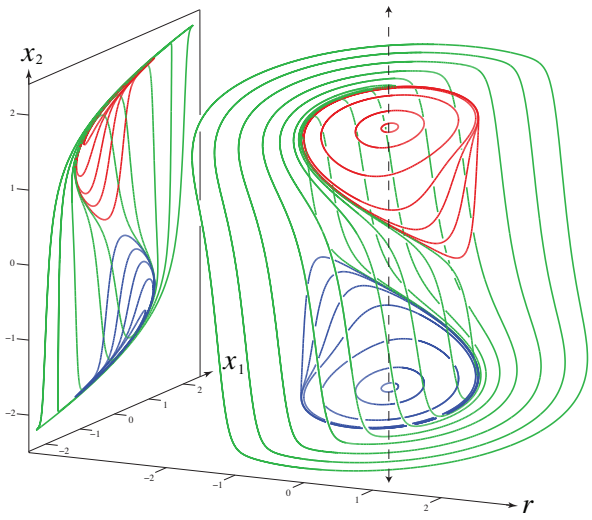

Figure 7: A phase portrait for the total dynamical system (4) and its projection to the agent state space. Only the orbits of the attracting set for ϕ_τ are shown. Every orbit of ϕ_τ is confined to an invariant cylinder and the collection of invariant cylinders are coaxial with the dotted vertical line.

orbit are bounded then it must accumulate to a periodic orbit. Thus every orbit of this type of system must either be periodic, accumulate to a periodic orbit, or be unbounded. If orbits are accumulating on both sides of a periodic orbit then the periodic orbit is an attractor for the total dynamical system restricted to the invariant cylinder containing the periodic orbit. In this example the orbits are periodic or accumulate to a periodic orbit.

Figure 7 shows a collection of periodic orbits for ϕ_τ. Each of these periodic orbits is an attractor for ϕ_τ restricted to the invariant cylinder containing the periodic orbit. However, technically, the periodic orbits are not attractors of ϕ_τ. This is because every neighborhood of a periodic orbit contains other periodic orbits and pairs of periodic orbits do not converge toward one another. It is only by restricting ϕ_τ to an invariant cylinder that there can be a neighborhood of a periodic orbit that contains only states which converge to the periodic orbit.

The total system, ϕ_τ, does have an attracting set though. This attracting set is the union of all of the periodic orbits which are attractors within their respective cylinders. This union of periodic orbits does have a neighborhood in which every orbit converges to some periodic orbit in the union. This attracting set is not an attractor because it does not contain an orbit which is dense inside of it. The projection by π to S_α of the periodic orbits in the attracting set are the attracting paths for this system. A collection of attracting paths are shown in Figure 7.

In this example there are three types of attracting paths. One type of attracting path are the hysteresis loops which are shown in green in Figure 6. The two other types of attracting paths are shown in red and blue. Paths like these occur in many systems exhibiting hysteresis but they are less commonly noticed.

Which specific attracting path is followed depends on the amplitude of oscillation for the input to the open dynamical system and the initial state. Hysteresis loops occur when the amplitude of the input oscillation is large. When the amplitude of the oscillating input is decreased below a critical value the hysteresis loops cease to exist and the two other attracting paths come into existence. As the amplitude continues to decrease to zero the red and blue attracting paths contract to a pair of points with coordinates $(0, \pm\sqrt{3})$ in S_α.

There are four trapped isotypes in this example (not shown in Figure 7). The four trapped isotypes corresponds to four different ways of approaching the attracting isotypes. One trapped isotype corresponds to approaching a green or red attracting path from "above" (*i.e.* from large positive values of x_2). A second isotype corresponds to approaching a green or blue attracting path from "below" (*i.e.* from large negative values of x_2). A third isotype corresponds to to approaching a red attracting path from "below". A fourth isotype corresponds to approaching a blue attracting path from "above".

3 Neural networks open to an environment

Neural networks are a broad class of models used in cognitive science, which are sometimes used to model actual circuits in animal brains, but are also used to model psychological processes in a biologically plausible way (see Rumelhart *et al.* [1986]). They are made up of interconnected processing elements or "nodes" which correspond to biological neurons. The connections between nodes ("weights") correspond to the function of synapses in transmitting information between biological neurons. In this section we show how the basic machinery of neural networks can be understood from the standpoint of open dynamical systems. Neural networks will be treated as agent systems ϕ_α with their own intrinsic dynamics, and a model environment ϕ_ϵ will be defined, as well as a way of coupling this environment to the neural networks we consider. We also describe how existing treatments of the concept of a mental representation can be understood in this framework.

3.1 Dynamical systems for the neural net

One commonly distinguishes between the set of possible patterns of activity across the nodes of a neural network, and the set of possible weights between nodes. The former is sometimes called an "activation space," the latter a "weight space" (*e.g.*, Churchland & Sejnowski [1992]). The full state space of the neural network is the Cartesian product of the activation space and the weight space. Following standard practice, we treat the weights as parameters which can vary (during learning) but which are otherwise fixed as the network changes state over time.

Let there be N nodes in the neural net, then $S_\alpha = \mathbf{R}^N$ is the state space for the neural net. The neural network's update rules define the dynamical system $\phi_\alpha : \mathbf{R}^N \times \mathbf{R} \to \mathbf{R}^N$. A specific instance of this type of model, the continuous Hopfield network, is described in the next section.

3.2 Representational structures in neural networks

A representation in a neural network is typically assumed to be a pattern of activity across its nodes or some subset of its nodes. An organism's representational repertoire–the set of representations it has acquired–is thus a set of patterns of activity which can be associated with a set of points in its network's activation space (*i.e.* a set of points in S_α). The merit of this approach is that it has made it possible to visualize many aspects of a neural network's behavior. Nearby points correspond to representations of similar objects, acquisition of conceptual structure corresponds to a partitioning of the state space into regions which represent different categories, and learning corresponds to a process of updating the weights so as to produce this partition (see Churchland & Sejnowski [1992]; Smolensky [1988]).

In philosophical discussions of representation, the states of an agent are typically associated with the objects they represent as follows: a state s of a cognitive system is said to represent an object o when s occurs in the presence of o.[11] This simple analysis can be used to analyze representational structures in neural networks (*cf.* the references to Churchland & Sejnowski [1992]; Smolensky [1988] above), and in separate work we have used this framework to analyze paths in very simple neural networks. However, in practice, more needs to be said about the relationship between physical process in the world and corresponding representational processes in an agent.

In particular, there is a complex relationship between the timing of neural processes and the timing of the mental states these give rise to (the points made here are reviewed in Koch [2004], ch. 15; also see Spivey [2006], and Smolensky

[11] For discussion of various approaches to this problem and the technical difficulties they encounter (mostly concerning mis-representation, *i.e.* cases when s occurs absent o), see Cohen [2004]. More general questions can be raised about the very notion of an object. There is a tradition going back at least to Kant in 1787 (Kant [1999]) which says that agents impose "object" categories on a world which has no such inherent divisions.

[1988], proposition 16). For example, there is a roughly 250 ms. time-lag between the occurrence of a retinal stimulus and a visual perception. The last 100 ms. of this process is thought to correspond to the time it takes for neural activity in the visual cortex to achieve a level of stability and coherence sufficient for normal visual experience. If retinal inputs are presented too briefly, or in the wrong sort of context, cortical activity will not achieve this level of stability, and the resulting visual experience may not accurately represent the environment. In the next section we consider a model which illustrates these points.

One novel aspect of our approach to representation is the fact that, by emphasizing paths in an agent space, we consider not just individual representational states, but representational *processes*. Moreover, we show how geometrical and topological tools can be used to analyze these processes.

3.3 Dynamical system for the environment

From the standpoint of a neural network, an environment is a means of producing inputs to a subset of its nodes, and of processing outputs from another subset of its nodes. To produce plausible inputs, it is useful to have a model of the environment. We make the simplifying assumption that the agent has a fixed position in the environment and that it is the objects which can change position. The objects will be represented by point particles. The level of stimulation produced by an object will depend only on the distance between the object and the agent. Consequently, for illustrative purposes, it will be enough for the space which contains the objects to be one dimensional. To keep the environment finite in size we will assume the space containing the agent and two objects is a circle. The circle will be parameterized by angle measured in degrees. The position of the agent will be 300^o (see Figure 8).

The state space for the environment is the configuration space for two distinguishable particles moving in the circle. This space is a 2-torus. The position of the m^{th} object will be denoted by r_m. The configuration of the two objects can be written as the position vector $\mathbf{r} = (r_1, r_2)$.

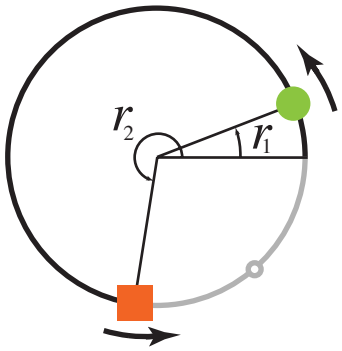

Figure 8: Schematic of the circle world. The position of the agent is fixed on the circle at 300^o (shown as an open gray circle) and its field of view extends 60^o in both directions (shown as gray arcs). The objects travel along the circle with object 1 (symbolized by the green disk) located at r_1 and object 2 (symbolized by the red square) located at r_2. At the moment shown object 1 has just left the agent's field of view while object 2 is just entering the agent's field of view.

Each object can move at its own velocity which we will usually assume to be constant. We denote the velocity of the m^{th} object by v_m. The velocities of the objects together forms a velocity vector $\mathbf{v} = (v_1, v_2)$. This gives us a dynamical system on S_ϵ.

$$\phi_\epsilon(\mathbf{r}, t) = \mathbf{r} + t\mathbf{v} \mod 360^o$$

Every dynamical system of this type can be continuously deformed to the identity dynamical system simply by contracting $\mathbf{v}$ to $(0, 0)$. The behavior of this type of dynamical system is fairly simple. When the ratio $v_1 : v_2$ is rational the orbits are periodic *i.e.* closed loops. When the ratio $v_1 : v_2$ is irrational the orbits are quasiperiodic. Quasiperiodic orbits are not strictly periodic but for any small number $\delta > 0$ there is an infinite sequence of future time intervals during which the orbit returns within a distance of δ from the initial condition and then leaves again.

We will focus on one particular type of dynamical systems for the environment in which both objects travel at the same speed. Both objects come into and out of view for the agent. We

call this the "diagonal world" because its phase portrait consists of parallel orbits which are diagonal with respect to the coordinate system for the torus (see Figure 12). We will also briefly consider the effect of an accelerating object.

3.4 Coupling between the neural network and environment

For simplicity we assume a one-way coupling between the environment and the neural network. The network receives inputs from the environment but does not act on that environment. Again, two-way couplings can be analyzed in an open dynamical systems framework which we will pursue in a later study.

The inputs are determined by the position of the two objects which travel about the circular world. Objects can have a differential impact on the sensory receptors of an organism. To model this we associate a *stimulus vector*, $\mathbf{u}_m$, to each object in our model environment. The stimulus vector can be thought of as the object's potential impact on the organism's sensory receptors under ideal conditions.

Each stimulus vector will be scaled based on how far the object is from the agent and these scaled stimulus vectors will be summed together to produce the vector of inputs to the neural network. The amount of scaling is generally a positive decreasing function, f, of the distance between the organism and the objects. We will use a piecewise linear scaling function:

$$f(d) = \begin{cases} 1 - d/60 & \text{if} \quad 0 \leq d \leq 60 \\ 0 & \text{else} \end{cases} \tag{5}$$

We call this a "scaling function". The strength of the input will be 1 if the corresponding object is at the same position as the agent and the strength of the input falls linearly to 0 as the object moves to 60^o away from the agent. Once the object is beyond 60^o away from the agent the strength of the input remains at 0.

If we let $d(r_m)$ be the distance between the agent (located at 300^o) and the m^{th} object then the effect of the environment is to add

$$f(d(r_1))\mathbf{u}_1 + f(d(r_2))\mathbf{u}_2$$

to the input nodes of the neural net.

4 The continuous two node Hopfield network

In this section we will analyze a specific neural network, a continuous Hopfield network, as an open dynamical system. Hopfield networks are recurrent neural networks which means that the output of node can exert an influence through the network that ends up influencing the input back to same node. Hopfield networks are interesting as open dynamical systems because their internal dynamics are relevant in shaping the paths which are followed in the network's state space. Because of their recurrent connections, Hopfield networks can respond to the same environmental input in many different ways depending on their own state.

Hopfield networks are sometimes referred to as "attractor networks" because when they are trained they have multiple stable fixed point attractors each of which is a "memory" that the network will settle into in response to a stimulus. We will see that, although they are known for their simplicity, Hopfield networks behave in extremely complex ways even when exposed to simple environments.

We begin this section by describing Hopfield networks and formally analyzing them as open dynamical systems. Then we give a traditional fixed point analysis of the closed system, followed by an analysis of the bifurcations of the closed systems. With these tools in place, we give an analysis of the paths which occur when the system is open to the diagonal world (defined in section 3.3). Though we do not classify the isotypes in this case, we do identify attracting paths with significant topologies, which correspond to documented psychological processes. Hence even with the very simple structure of an attractor network subjected to moving objects we observe known psychological phenomena.

4.1 Basic description

Hopfield networks consist of identical copies of a single type of neuron. The neurons can have either discrete or continuous values. In this text we will use continuous valued neurons. The neu-

ron's values are allowed to be any real number but the dynamics tends to keep the values around ± 1. The state of each neuron is determined by the inputs it receives from the environment and from every other neuron. For continuous valued neurons each neuron also has a self connection.

The state space for the Hopfield network with N neurons is $\mathbf{R}^N$ and we denote the state of the neurons by $(x_1, \ldots, x_N)$. The dynamics of a continuous Hopfield network which is not coupled to the environment 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 $g(x_i) = (2/\pi)\tan^{-1}(\pi\lambda x_i/2)$ is shaped liked the sigmoidal function and λ is a parameter which determines the steepness of g. Following Hopfield's example (Hopfield [1984]) we will let $\lambda = 1.4$. These differential equations resemble the Hodgkin/Huxley model for a biological neuron (Hodgkin & Huxley [1952]). The R_j can be thought of as the electrical resistance of a biological neuron's membrane and for convenience we let $R_j = 1$ for all j.

Hopfield networks are trained to recognize objects by using a Hebbian learning rule *i.e.* the connections between neurons which should be in the same state when recognizing an object are strengthened. More precisely we choose state vectors from $\{-1, 1\}^N$ to represent objects. Suppose we have n objects and let $\xi_j = (\xi_{j1}, \ldots, \xi_{jN}) \in \{-1, 1\}^N$ be the state vector we choose for the j^{th} object. The weights of the Hopfield network are then given by

$$w_{ji} = \frac{1}{N}\sum_{k=1}^{N} \xi_{jk}\xi_{ki}$$

One consequence of this type of learning rule is that the weights are symmetric: that is, $w_{ji} = w_{ij}$ for all i, j.

Once trained, Hopfield networks share with biological neural networks a capacity for associative memory recall. Even when the stimulus received by a Hopfield network differs slightly from that produced by an object it has been trained to recognize, the network can still recognize the object. One way to understand how a Hopfield network does this is by a fixed point analysis. For the neural network the act of recognizing an object is the process of settling down to an attracting fixed point. When a stimulus resembles an object the network has been trained to recognize the network is driven into the basin of attraction for the fixed point and subsequently moves towards that fixed point (see the phase portrait in Figure 10, which corresponds to a network with two memories).

In fact, however, the actual position of the fixed points in the state space of a Hopfield network varies as the inputs change and they can even come in to and go out of existence. Consequently a Hopfield network doesn't actually enter into the state that it is trained to reach when the object is present but rather a state which is close to the desired state. It is possible for the network's state to remain as closely as desired to a fixed point as the inputs change so long as the inputs change slowly enough. But if the inputs change rapidly the position of the fixed points in the state space is only a rough guide to the state of the neural network and if a fixed point bifurcates out of existence it can no longer be used at all.

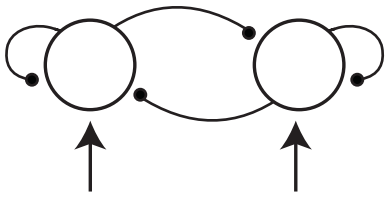

Figure 9: Topology of the two node Hopfield network.

To illustrate the basic principles it is sufficient to consider (as Hopfield did in Hopfield [1984]) the two neuron case (see Figure 9). In this example we train our Hopfield network to be a two object detector. We want the network to enter the state $\xi_1 = (1, -1)$ when it detects object 1 and to enter the state $\xi_2 = (-1, 1)$ when it detects object 2. This gives $w_{12} = w_{21} = -1$. Our agent system, ϕ_α is then given by the solu-

tions to the ODE:

$$\dot{x}_1 = -x_1 - \frac{2}{\pi}\tan^{-1}(0.7\pi x_2)$$
$$\dot{x}_2 = -x_2 - \frac{2}{\pi}\tan^{-1}(0.7\pi x_1).$$

The phase portrait for the decoupled Hopfield network is shown in Figure 10. When the Hopfield network is embedded in an environment we will see its behavior to be related to but substantially different from the decoupled network. In particular, as the inputs to the network change, the attracting fixed points move, and the paths follow these moving fixed points. Notice that the diagonal $x_1 = x_2$ is the boundary of the basin of attraction for the fixed points. Initial conditions on either side of the diagonal will approach one or the other fixed point attractor. We will think of the diagonal as dividing the agent state space into two representational regions, corresponding to representations of the first and second object, respectively.

The total state space for the environment and neural network is the Cartesian product of the state space for the environment and the state space for the neural network *i.e.* $S_\tau = \mathbf{T}^2 \times \mathbf{R}^2$. The map π will be the projection of S_τ onto $\mathbf{R}^2$. The dynamical system φ_τ will be the product dynamical system $\iota_\epsilon \times \phi_\alpha$. For simplicity object 1 will have stimulus vector $(1, 0)$ and object 2 will have stimulus vector $(0, 1)$. The total dynamical system $\phi_\tau : (\mathbf{T}^2 \times \mathbf{R}^2) \times \mathbf{R} \to (\mathbf{T}^2 \times \mathbf{R}^2)$ will be given by solutions to the ODE:

$$\begin{aligned}
\dot{r}_1 &= v_1 \\
\dot{r}_2 &= v_2 \\
\dot{x}_1 &= -x_1 - \frac{2}{\pi}\tan^{-1}(0.7\pi x_2) + f(d(r_1)) \\
\dot{x}_2 &= -x_2 - \frac{2}{\pi}\tan^{-1}(0.7\pi x_1) + f(d(r_2))
\end{aligned}$$

The continuous deformation between ϕ_τ and φ_τ is obtained by letting (v_1, v_2) go to $(0, 0)$. This completes our specification of a Hopfield network as an open dynamical system.

4.2 Bifurcation analysis

Hopfield networks were originally designed to work with their input levels held at constant values. When the inputs to a Hopfield network are held constant they can be regarded as parameters of a dynamical system which governs the network. This type of traditional analysis will be useful to our subsequent analysis of the Hopfield network when it is embedded in an environment.

Specifically we treat the input levels, $(I_1, I_2) = f(d(r_1)), f(d(r_2))$ as the parameters of the dynamical system for the dynamical system given by the ODE:

$$\dot{x}_1 = -x_1 - \frac{2}{\pi}\tan^{-1}(0.7\pi x_2) + I_1$$
$$\dot{x}_2 = -x_2 - \frac{2}{\pi}\tan^{-1}(0.7\pi x_1) + I_2 \qquad (6)$$

Depending on the input, there is either one fixed point, three fixed points, or else the network is at a bifurcation point between these two cases.

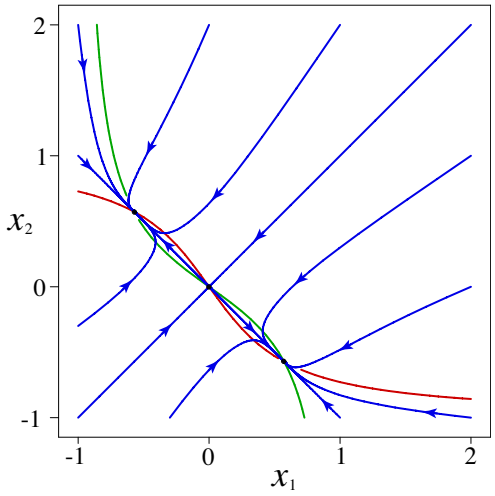

Figure 10: A phase portrait for the decoupled Hopfield network, ϕ_α. The nullcline for $\dot{x}_1 = 0$ is shown in green and the nullcline for $\dot{x}_2 = 0$ is shown in red. The orbits are shown in blue.

The bifurcation analysis for this system is facilitated by the use of nullclines. The nullcline for x_j is the locus of points where the rate of change of x_j is zero[12]. At a fixed point the value of (x_1, x_2) does not change and $(\dot{x}_1, \dot{x}_2) = (0, 0)$. So the fixed points are the intersection of the nullclines for x_1 and x_2. The equations for the

[12]The black curve in Figure 4 is an example of a nullcline for ϕ_α in the hysteresis example.

nullclines can be found by setting $\dot{x}_1 = 0$ and $\dot{x}_2 = 0$ into the ODE 6. After a little rearrangement the equations for the nullclines become

$$x_1 = I_1 - g(x_2) = I_1 - \frac{2}{\pi}\tan^{-1}(0.7\pi x_2)$$
$$x_2 = I_2 - g(x_1) = I_2 - \frac{2}{\pi}\tan^{-1}(0.7\pi x_1)$$

Each of these is a curve with a sigmoidal shape. Changing the value of I_1 or I_2 simply translates the nullclines without changing their shape. The nullcline for x_1 has vertical asymptotes at $x_1 = I_1 \pm 1$ and the nullcline for x_2 has horizontal asymptotes at $x_2 = I_2 \pm 1$. The fixed points must therefore lie inside the square $[I_1 - 1, I_1 + 1] \times [I_2 - 1, I_2 + 1]$. When the absolute values of I_1, I_2 are large only the asymptotic portions of the nullclines lie inside this square and so there is only a single fixed point and it is attracting.

As I_1, I_2 move close to zero fixed points bifurcate into existence. For a bifurcation value of (I_1, I_2) the nullclines are tangent to each other at the intersection point. This fact can be used to determine the relationship which must hold between I_1 and I_2 at a bifurcation.

The tangent vectors to the nullclines can be computed from a parametric form for the nullclines.

$$\begin{array}{lll} (I_1 - g(x_2), x_2) & \text{has tangent} & (-g'(x_2), 1) \\ (x1, I_2 - g(x_1)) & \text{has tangent} & (1, -g'(x_1)) \end{array}$$

The tangents are parallel when their cross product is the zero vector. This gives

$$\det \begin{pmatrix} -g'(x_2) & 1 & 0 \\ 1 & -g'(x_1) & 0 \\ \mathbf{i} & \mathbf{j} & \mathbf{k} \end{pmatrix}$$

$$= \begin{pmatrix} 0 \\ 0 \\ g'(x_1)g'(x_2) - 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

Since

$$g'(x) = \frac{\lambda}{1 + \left(\frac{\pi}{2}\lambda x\right)^2}$$

is an algebraic function the equation $g'(x_1)g'(x_2) - 1 = 0$ gives an algebraic relationship that must hold between x_1, x_2 at a bifurcation. We can solve $g'(x_2) = 1/g'(x_1)$ for x_2 in terms of x_1 and write out parametric equations for the bifurcation values of I_1, I_2 in terms of x_1.

$$\begin{aligned} x_2(x_1) &= \pm\frac{2}{\pi\lambda}\sqrt{\frac{\lambda^2}{1 + \left(\frac{\pi}{2}\lambda x_1\right)^2} - 1} \\ I_1 &= x_1 + g(x_2(x_1)) \\ I_2 &= x_2(x_1) + g(x_1) \end{aligned}$$

These parametric equations define a simple closed curve which is symmetric under reflection about the diagonal $I_1 = I_2$ and which has two cusps that lie on the diagonal (see Figure 11).

When (I_1, I_2) is outside of the curve there is exactly one fixed point. As (I_1, I_2) move from the outside to the inside of the curve a saddle node bifurcation occurs[13]. We obtain three fixed points two of which are attracting when (I_1, I_2) is inside the bifurcation curve.

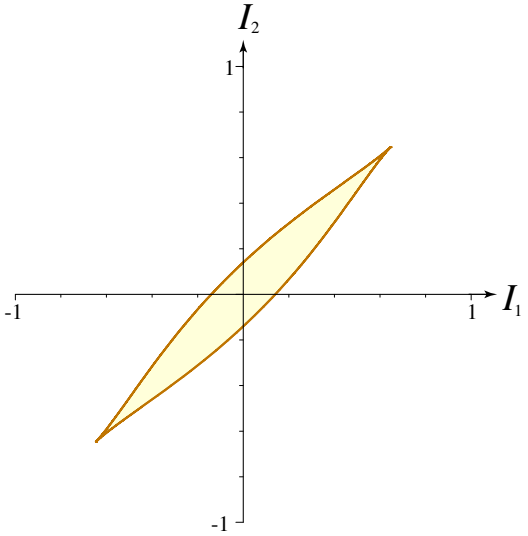

Figure 11: Classic bifurcation analysis of Hopfield network subject to fixed inputs. For (I_1, I_2) inside the shaded region there are three fixed points in the neural network's state space and for (I_1, I_2) outside the bifurcation curve there is just a single fixed point.

[13] Unless the transit is made through one of the cusps in which case a different type of bifurcation called a pitchfork bifurcation occurs.

4.3 Effect of the environment on the bifurcations of the network.

We now use the bifurcation analysis from the previous section to begin to study the effect of the diagonal world on the Hopfield network. To get a qualitative understanding of the impact of this environment on the neural network we determine the preimage of the bifurcation curve in the (I_1, I_2) input space under the map:

$$(r_1, r_2) \to (f(d(r_1)), f(d(r_2))) = (I_1, I_2)$$

This is straightforward since f is piecewise linear. The image of $f \times f$ is the square $[0,1]^2$ and only the portion of the bifurcation curve in the square has a preimage. The preimage of the bifurcation curve still turns out to be a simple closed curve in S_ϵ. This curve is shown in Figure 12 in dark brown. This curve partitions the state space of the environment into two regions each of which corresponds to a particular number of fixed points which would exist in the state space of the neural net if the environmental state were held fixed.

One of these regions is topologically an open disk and environmental states inside it produce three fixed points in the agent state space, two of which are attracting. These are environment states in which both objects are either out of view or barely in view by the agent.

The other region in the 2-torus has the topology of a two dimensional tubular neighborhood of a figure eight space. This region corresponds to one or both objects being near the agent. These states result in there being just a single fixed point in the neural net's state space and this fixed point is attracting.

In one loop of the figure eight shaped region the corresponding unique fixed point in the neural net state space is near $(1, -1)$ while in the other loop the corresponding unique fixed point in S_α is near $(-1, 1)$. Neural net states near these fixed points correspond to representations of one of the objects in the environment. Points in the figure eight shaped region where the two loops meet correspond to representations of both objects.

We will examine the way the orbits of the diagonal world impact the neural network (see Figure 12), which will involve fixed points of the agent system coming into and going out of existence as objects come into view and go out of view. However, it is worth noting that it is in fact possible for some environmental dynamical systems to have a periodic orbit in which objects come into and go out of view and in which a unique attracting fixed point of the agent system moves continuously back and forth from a neighborhood of $(1, -1)$ to a neighborhood of $(-1, 1)$ without any other fixed points coming into existence.[14]

4.4 The Hopfield network in the diagonal world

Having put some basic apparatus in place for understanding the Hopfield network, let us now open it to the simple environment described above (the "diagonal world.") The orbits of this environmental dynamical system are simple closed curves which can be pictured as a collection of diagonal lines, as in Figure 12.

We will see that the attracting sets for the total dynamical system are composed of periodic orbits like with the hysteresis example. The projection of the periodic orbits to the agent state space are closed curves that sometimes cross themselves.

In analyzing these paths we focus on (1) the constant speed of the objects and (2) the angular separation between the two objects. Note that the constant speed of the objects is a parameter for a family of total dynamical systems. On the other hand even though the angular separation between the objects is constant over time when both objects are moving with the same velocity it is not, strictly speaking, a parameter. The angular separation is an example of a dynami-

[14] As can be seen in Figure 12, it is possible to connect a state in one loop of the figure eight shaped region to a state in the other loop by a curve which lies entirely within the figure eight shaped region. Changing the environmental state continuously along such a curve would cause the corresponding unique fixed point in the neural net's state space to move continuously from being near $(1, -1)$ to being near $(-1, 1)$.

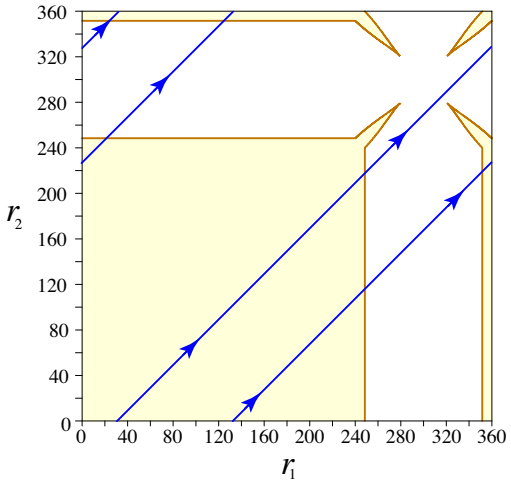

Figure 12: The state space for the environment, S_ϵ, which is a 2-torus (top and bottom edges are to be identified as well as the left and right edges). The preimage in S_ϵ of the bifurcation curve is shown in dark brown. This curve is connected in one piece and it subdivides the torus into two regions. When the environmental state is in the larger of these two regions (shaded tan) the agent dynamical system has three fixed points. When the environmental state is in the other unshaded region (a figure eight wrapped around the torus) the agent dynamical system has one fixed point. The blue curves are two orbits of the diagonal world. The orbit with an angular separation of about 30^o crosses the bifurcation curve twice per period. The orbit with a angular separation of about 130^o crosses the bifurcation curve four times per period.

cal invariant *i.e.* a quantity which is a function of the state of the dynamical system and which is constant on each orbit. In these examples we can specify the orbits of the environment dynamical system with the angular separation. In the hysteresis example the distance from the x_2-axis was a dynamical invariant for that system and the cylinders coaxial about the x_2-axis were invariant sets.

In our Hopfield network examples the Cartesian product of a periodic orbit in S_ϵ with S_α is an invariant set of ϕ_τ. These invariant sets have the topology of $\mathbf{S} \times \mathbf{R}^2$ and they are also called cylinders. Of course it can be difficult to visualize the collection of these three dimensional cylinders in the four dimensional state space S_τ but since the Hopfield network does not affect the environment in these examples the angular separation between the objects is a dynamical invariant for the ϕ_τ which we can use to specify individual invariant cylinders in S_τ. Every orbit of ϕ_τ is contained in some invariant cylinder.

As can be seen in Figure 12 when the angular separation is small the environmental state spends most of its time in a region of S_ϵ which produces inputs that result in three fixed points in S_α and it spends little time in a region which produces inputs that result in a single fixed point in S_α. When the angular separation is large the amount of time the environmental state spends in these two regions is more equitable.

We will see that at slow speeds the angular separation between the objects produces a variety of looped paths in the agent space which can be interpreted as representational processes. As the speed of the moving objects is increased bifurcations occur for the restriction of ϕ_τ to the invariant cylinders which involve the attracting periodic orbits. Also the paths in the agent space become more tightly confined to a small region of S_α. The paths and the changes they go through are consistent with a variety of psychophysical phenomena.

At slow speeds the Hopfield network can be analyzed mathematically using a quasistatic variation of parameters approach. At fast speeds the technique called "averaging" will be useful. Its at intermediate speeds that the behavior is the most complicated and where the open dynamical systems approach becomes especially valuable.

To compare the behavior of the closed agent system with that of the open system, in the figures below a pair of red and green curves show the nullclines of the decoupled Hopfield network, whose intersections shows where the three fixed points of the decoupled system are. The diagonal line in these figures corresponds to the basin boundary in the decoupled agent system. Blue curves represent the paths the system follows when it is embedded in an environment.

4.5 Representational processes in the open Hopfield network

We now consider the representational processes entailed by the paths that occur in the open Hopfield network. We consider different angular separations between objects (so that the duration between the objects coming into view can be short or long) and different velocities for the objects. We will see that these differences in the environment are reflected in the corresponding representational processes. We will also see that the agent does not directly represent objects, but that there are significant (and psychologically plausible) differences between what happens in the environment and the way the agent represents the environment.

We separately consider the psychological relevance of three cases: (1) objects moving at slow velocities, (2) objects moving at fast velocities, and (3) the transitions that occur in the network at intermediate velocities. In each case we see that the resulting representational processes capture known psychological phenomena. Following standard interpretations of Hopfield networks, we assume that the attracting fixed points which the decoupled network is trained to approach upon exposure to particular objects, correspond to representations of those objects. And since the network generally never actually reaches these fixed points in finite time we also make the assumption (usually tacit in the Hopfield literature) that all states in some region around each of the attracting fixed points also correspond to representations of the relevant object.

Slow object speeds

When the objects are moving relatively slowly, the system has time to come close to an attracting fixed point of the closed system, though this "fixed" point moves as the inputs change. The path of the open dynamical system follows this moving "fixed point". Both the angular separation between the two objects and the specific value of the slow speed have an impact on the shape of the path. To give a sense of the relationship between the angular separation and the path-shape, consider Figure 13, which shows the paths which occur in the agent space when the speed is $1/2$ and the angular separations, $r_2 - r_1$, are 60^o, 90^o, 180^o, 270^o, and 300^o. In the middle figure, the case of 180^o separation, the objects are maximally spread apart so that the system has time to separately respond to each of the two objects. When the objects are closer to each other the path tends to stay in the region corresponding to the last object the agent sees before both go out of view.

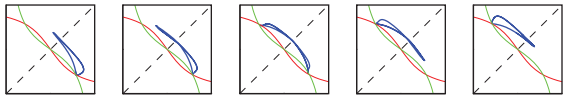

Figure 13: The agent's state space for the Hopfield network in the diagonal world where the objects move with speed $1/2$. The five views show what happens when the angular separation, $r_2 - r_1$, between the objects is 60^o, 90^o, 180^o, 270^o, and 300^o respectively. Each angular separation specifies an invariant cylinder and the attracting paths in the figure are projections of attracting periodic orbits within the specified invariant cylinder.

In this example, we assume that only those agent states that are visited with sufficiently low velocity will produce a representation of an individual object (this captures the idea, discussed in section 3.2, that only network states which are sufficiently stable produce normal visual experience).

We begin with the case of 180^o angular separation, which corresponds to the relatively everyday case of two objects coming separately into view. The agent first represents one object, then the other, as the two objects come into and go out of view. The Hopfield network follows a set of states which correspond to each object, as well as a series of transitional states. The attracting path followed by the neural network (shown in 14) forms a figure eight shaped curve in S_α, where the outer portion of each loop of the figure eight corresponds to seeing one of the objects, and the middle portion corresponds to transitional states. The labels in Figure 14 show what

happens when the first object comes into and goes out of view.

Notice that there is a time lag between what occurs in the environment and what the agent represents. For example, in Figure 14, when object 2 (the red square) comes in view the agent is still perceiving object 1 (the green disk). This captures the idea that there is a time-lag between initial presentation of an object to an agent and the agent's representation of that object. [15]

The rate of change of the agent's state tends to vary along the attracting path because of the intrinsic dynamics of the Hopfield network. When an object first comes in to view the system is pulled rapidly to the other side of the agent space, but the system slows down by the time the object is nearby. The transitions between objects occur in a rapid manner compared with the duration in which the objects are in view for the agent.

We have assumed that only those network states which are changing at a low velocity produce representations of particular objects. This raises the question of how we should interpret the higher velocity states which occur between these representations. We have left open the possibility that some other form of representational process might occur. In fact, the idea that there are intermediate forms of representation between stable representations of individual objects has a long history. According to William James, "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). Interestingly, these hypothesized transitory phases of perception have until recently received comparatively little study relative to stable phases, because of the difficulty of studying transitory phases. However, recently some researchers have begun to introduce techniques for experimentally studying these more complex periods of change in experience (see Spivey [2006]).

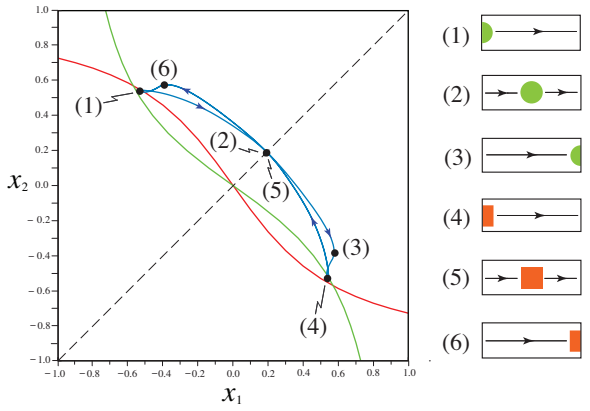

Figure 14: The attracting path of the Hopfield network coupled with the diagonal world with speed 1/2 and angular separation of 180^o. In the figures on the right we see the location of the objects in the agent's field of view for the labeled moments. Notice that the objects location does not directly correspond to what the agent is perceiving: there is a time lag between an object entering the agent's field of view, and the agent perceiving the object.

When the objects are closer to each other (but more than 40^o apart) the attracting path is no longer symmetrical and most of its length is in the representational region of whichever object is the closest behind the other. For example consider the diagonal world with speed 1/2 and angular separation of 60^o with object 1 following behind object 2. The path is largely in the region corresponding to object 1 (see Figure 15). When object 2 comes in to view the path is pulled towards the object 2 region. But soon after object 2 comes in view, object 1 comes in to view, and as object 2 goes out of view object 1 dominates

[15] This model also implies, implausibly, that if no object is in view, that the system will settle in to the fixed point attractor nearest the state it was in when the last object was in view. For example, if the agent recently perceived a green disk, which goes out of view, then the model implies that the agent will continue to perceive a green disk. While there is some plausibility to this over very brief periods, it is obviously implausible over longer time periods. To capture the idea that a system stops representing anything in the absence of inputs we could add a decay term to x_1 and x_2 which would eventually take the system to the origin, a state which we can regard as representing no objects.

the system, pulling it back towards the object 1 region. When object 1 goes out of view the system remains very near the closed system attractor corresponding to object 1. The network still goes through a sequence of four saddle node bifurcations as the two objects come into and go out of view but now the timing of the bifurcations causes the system to spend more time representing object 1.

The path shape observed here is consistent with a perceptual phenomenon known as masking, which (in the visual case) "occurs whenever the visibility of one stimulus, called the target, is reduced by the presence of another stimulus, designated as the mask" (Breitmeyer & Ogmen [2000], p. 1572). Even though two distinct stimuli occur, only one is perceived. More specifically, the path shape illustrates "backward" masking, where the second object to appear (the mask) reduces the visibility of the first (the target). Backward Masking has been observed in a wide variety of conditions involving visual and acoustic stimuli. Here we see that a very simple attractor network coupled to periodically changing inputs can produce masking effects, where (in the example above), object 1 is the target, and object 2 is the mask.

Fast object speeds

We now consider what happens when both objects move at fast speeds. At fast speeds we can use a technique known in dynamical systems theory as averaging to approximate the behavior of the neural network. This is like clamping the inputs at the average value the two nodes receive. This is easily computed when the speed is constant.

$$\overline{I}_1 = \frac{1}{T}\int_0^T f(d(r_1(t)))dt = \frac{\frac{1}{2}\cdot 120^o \cdot 1}{360^o} = \frac{1}{6}$$
$$\overline{I}_2 = \frac{1}{T}\int_0^T f(d(r_2(t)))dt = \frac{\frac{1}{2}\cdot 120^o \cdot 1}{360^o} = \frac{1}{6}$$

where T is the common period of r_1, r_2. The

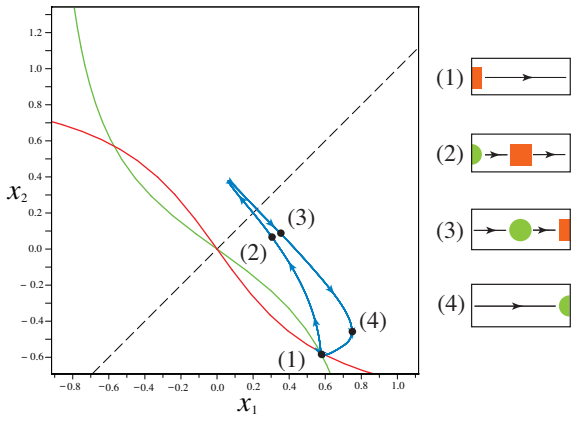

Figure 15: The attracting path of the Hopfield network coupled to diagonal world at velocity 1/2 and angular separations of 60^o. The figure illustrates a process whereby (1) object 2 (the red square) comes into view, (2) object 2 passes the agent while object 1 (the green disk) comes into view, (3) object 1 passes the agent as object 2 goes out of view, (4) object 1 goes out of view. Note that the path does not enter the representational region for object 2, even though object 2 comes into view. Object 1 "masks" object 2, so that only object 1 is perceived.

ODE for our averaged system is then

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

There are three fixed points in S_α for the dynamical system which solves this ODE. There is a saddle point on the diagonal $x_1 = x_2$ and two attracting fixed points symmetrically placed about the diagonal. The angular separation between the objects has little impact on the neural net. The neural network's state travels to whichever attracting fixed point is closest to the initial condition.

Technically averaging only provides an approximation. In actuality the neural network's state can follow one of two small tightly curved paths around either of the attracting fixed points of the averaged system. The larger the speed the more tightly the paths are confined to the attracting fixed points of the averaged system (see

Figure 16).

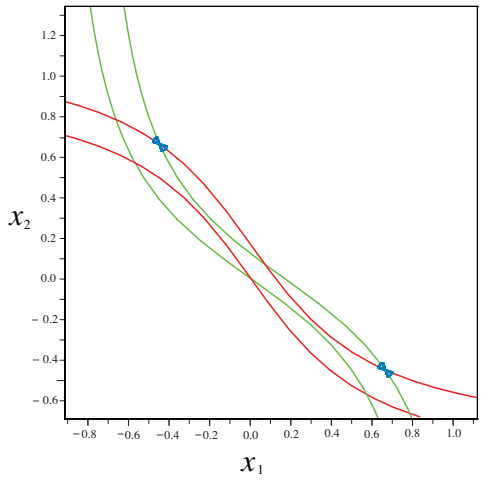

Figure 16: The attracting paths for the Hopfield network in the diagonal world with speeds $v_1 = v_2 = 10$. The nullclines for a system with inputs clamped at average values are also shown along with the nullclines for the closed agent system. Notice that the attracting paths are two small figure eights centered on the fixed points of the averaged system.

The high velocity case produces rapidly oscillating inputs to the agent's sensors. The agent's representation of the oscillating inputs is ambiguous, much like the famous ambiguous figures or "bistable percepts" of Gestalt psychology (the Necker Cube, the face-vase, etc.). While those ambiguities arise from stable inputs; it is known that the same type of ambiguities can also arise from rapidly oscillating stimuli such as those that occur in this example (see Pressnitzer & Hupé [2006]). Moreover, which of the two paths the system follows depends on which state the system began in. This is consistent with the experimental literature on ambiguous figures (reviewed in Long *et al.* [1992]), according to which a brief prior exposure to an unambiguous figure "primes" or "sets" the perceptual system, so that when subjects are subsequently exposed to the ambiguous figure they are more likely to perceive the one for which they have been primed. We can view the initial state of the agent as corresponding to prior exposure of a priming stimulus, which biased the system to interpret the ambiguous figure one way or another.

Transition from slow speed to fast speed

We now consider the changes that occur when the network transitions from very slow to fast speeds. When the objects are moving very slowly, with a speed near 0, the system has time to come very close to an attracting fixed point of the closed system, though this "fixed" point moves as the inputs change. The path of the open dynamical system follows this moving "fixed point". This is a quasistatic variation of parameters type argument. When the objects are moving very quickly we have seen that the attracting paths stay near whichever state the neural network began in.

We can use these facts to infer the existence of a bifurcation which occurs in the agent space when the velocity is increased. We first note that the figure eights which occur at high velocities are the projection of attracting periodic orbits in S_τ of the total dynamical system, ϕ_τ. Although it is mathematically possible for an orbit which is not periodic to project to a figure eight it would be an exceptional circumstance. Second, we note that the appearance of a second figure eight in S_α as the speed increases suggests that a bifurcation of periodic orbits occurs which is analogous to the saddle node bifurcation of fixed points. The bifurcations can happen within the invariant cylinders of S_τ and the periodic orbits are projected down to closed paths in S_α. This is consistent with the numerical analysis of the total system shown in 17. With the appearance of a new attracting periodic orbit for the restriction of ϕ_τ to an invariant cylinder there should also be a new repelling periodic orbit nearby but precisely because it is repelling it is difficult to detect numerically.

Sample bifurcations are pictured in Figure 17 for several angular separations, at speeds just below and above the bifurcation values. The bifurcations of the periodic orbits appears to occur along a curve which is the graph of a unimodal function for the object's speed in terms of angular separation. The bifurcation value appears to

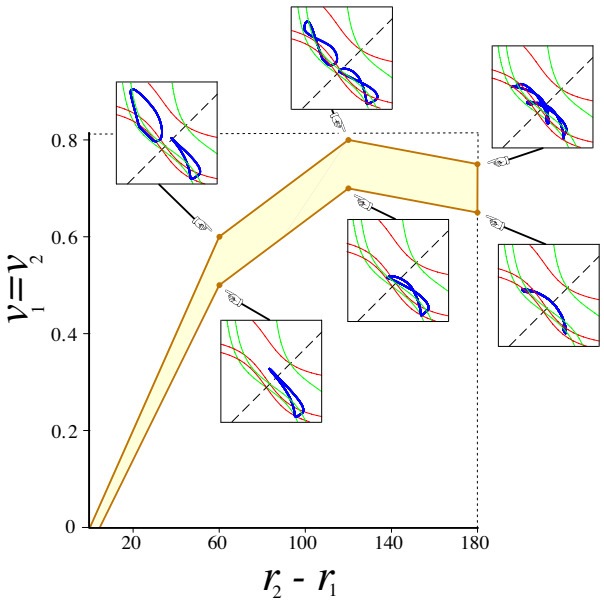

Figure 17: The restriction of ϕ_τ to an invariant cylinder (specified by the angular separation $r_2 - r_1$) appears to undergo a saddle node type of bifurcation with an attracting periodic orbit as the common speed, $v_1 = v_2$, of the objects is varied. Each inset shows the attracting paths in S_α. The bifurcation value of the speed varies from one invariant cylinder to another and has only been estimated to be within the shaded region.

decrease rapidly to zero as the angular separation between the objects goes to zero. For small angular separations between the objects the new attracting path which first emerges after the bifurcation is not a figure eight but a simple closed curve.

Which of the two attracting paths the agent follows when the objects' speed is above the bifurcation value depends on the initial conditions of the total system, which involves both the agent's state and the environmental state.[16]

These bifurcations are consistent with a broad range of observed perceptual phenomena. In many contexts qualitative changes occur as the temporal frequency of a pair of alternating stimuli is changed. For example, in certain conditions, when two stimuli are presented to a subject in increasingly rapid sequence, they are initially perceived as separate objects but beyond a threshold are perceived as a single object in motion. This is "apparent" or "stroboscopic" motion, whose experimental study dates back to Wertheimer [1912]. Similarly with tones: if tones are presented at a wide enough temporal interval they will simply be heard as individual tones. However, when the are presented in sufficiently rapid succession, they group and fuse together in various ways, often forming separate streams in an "auditory scene" (see Bregman [1990]). In some cases these fused percepts are ambiguous between several possible interpretations. For example, when a sequence of tones A and B are presented to a subject at a low frequency, they are heard as A's and B's in an unambiguous slow sequence. However, when they the frequency of presentation is increased, the oscillating stimulus becomes ambiguous: it will either be heard as one stream of "galloping" ABA's or as two streams A-A-A ... and B-B-B ... (Pressnitzer & Hupé [2006], p. 1351). Though priming effects have not been studied for these stimuli, the prevalence of priming in multistable perception suggests that these effects would be present in this case, so that which way the sequence would be heard would depend on the prior state of the agent.

Mixed velocities

It is interesting to consider slightly more complicated dynamical systems for the environment because their impact on the paths that occur in the agent space can be dramatic. We will not give an analysis of the representational processes implied by these cases, but given that such inputs occur in natural environments, they could be relevant to future cognitive research.

When the two objects move at different fixed speeds that are commensurate the orbits for the environmental dynamical system are parallel closed curves in the torus. The two objects moving at

[16] It turns out to to be hard to determine from the initial conditions which attracting path the agent will follow. In fact it is possible for the agent to go to either attracting path from the same starting state in the agent space and from the same orbit of the diagonal world.

the same speed can be regarded as a special case of moving with commensurate speeds. Just as with the same speed examples when the speeds are commensurate, in this case there are two attracting periodic orbits in the total system for each orbit of the environmental dynamical system. Which of these two periodic orbits the total system tends to depend on the initial condition.

The projection of these periodic orbits into the neural network's state space resembles stretched out versions of Lissajous figure (see Figure 18). As the speeds of the objects are increased without changing their ratio the Lissajous figures contract to the pair of attracting fixed points of the averaged system. The relative periods of the stretched Lissajous figures does not change but the amount of distortion decreases and they tend to become more like classical Lissajous figures (see Figure 19).

Classic Lissajous figures can be thought of as the projection of simple closed curves in a symmetric 2-torus with zero Gaussian curvature. Such a torus cannot be isometrically embedded in a three dimensional Euclidean space but it can be isometrically embedded in S_τ. The numerical evidence for the Hopfield network suggests that the state space for the full system contains two invariant tori which contain the attracting periodic orbits of the restriction of ϕ_τ to the invariant cylinders. As the objects speed increases the invariant 2-tori become smaller and appear closer in form to a symmetric 2-torus with zero curvature.

At high speed the attracting sets for ϕ_τ are likely to be a pair of symmetric 2-tori. Numerical evidence indicates that as the speed is lowered the attracting sets become, larger, less symmetrical, and eventually one of the tori is pinched off into a sphere.

When the velocities of the objects are not commensurate the dynamical system for the environment is quasiperiodic. Numerical evidence indicates that the full system is quasiperiodic as well with orbits dense inside of each of two tori (which would make them attractors).

When the object's velocity is not fixed more complicated types of behavior can occur and it appears that acceleration of an object can produce a quasiperiodic transition to chaos in the Hopfield network (see Figure 20).

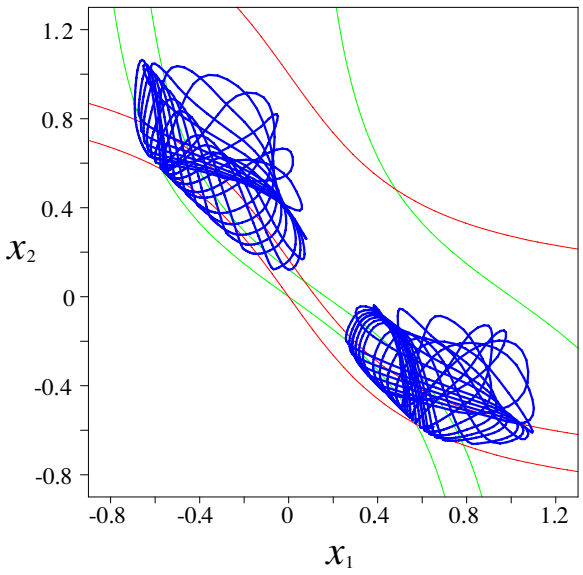

Figure 18: The attracting paths when the objects velocity's are in a 20:13 ratio. They resemble two distorted Lissajous figures centered on the fixed points of the averaged system. The nullclines for a system with inputs clamped at average values are also shown along with the nullclines for the closed agent system and the nullclines corresponding to both inputs clamped at their maximal values.

4.6 Generalization to multiple objects

Our discussion of Hopfield networks focused on the case where a two-node network has been trained to recognize two distinct objects, so that the state space of the closed system has two fixed point attractors, one corresponding to each object. The basins of attraction of these two fixed points correspond to "representational regions" in the network's state space. We have seen that the way the network visits these two regions is strongly affected by the nature of the environmental inputs to the system. As the velocity of the objects in the networks' environment changes, and as the angular separation (and hence, temporal interval) between them is altered, a wide array of representational processes are observed, including "flights and perches" relative to stimu-

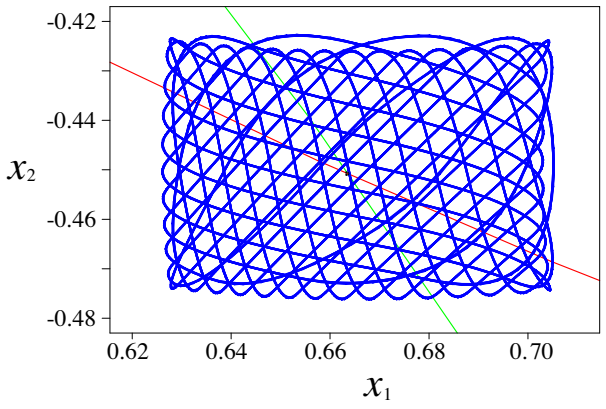

Figure 19: An attracting path occupies a rectangular shaped region and resembles a classic Lissajous figure. The objects speeds are in the same ratio as in Figure 18 but twelve times faster. The path's size has been reduced by roughly one twelfth.

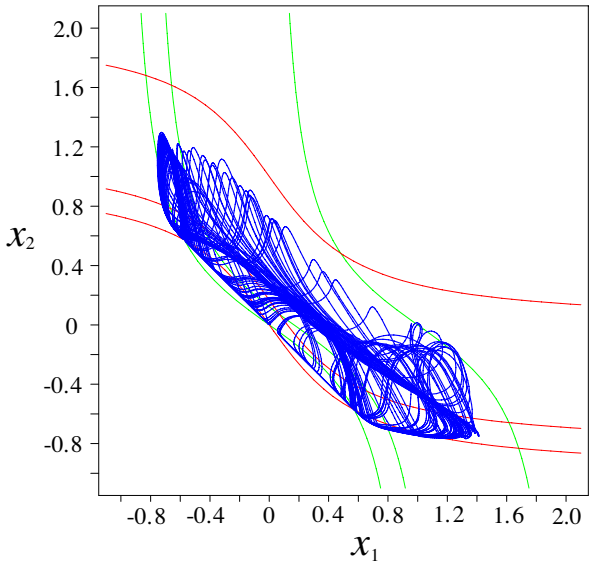

Figure 20: The attracting path in which object 1 travels at a constant velocity and object 2 travels at a constant acceleration. The path resembles the projection of a chaotic attractor.

lus presentations and transitions, masking of one object by another, qualitative perceptual changes as temporal frequency increases, and ambiguous representations biased by previous activity.

These observations generalize from the two object/two node network case to larger networks in worlds containing $n > 2$ objects. The agent systems in such cases (assuming a trained network with sufficiently many nodes) contain n basins of attraction each of which contains its own representational region. In such cases we expect that analogous behavior will occur. For example in the Hopfield network trained to recognize five objects coupled to a circle world containing those five objects, at slow-velocity and with sufficient distance between objects (and a narrow enough field of view) the path will be a closed curve passing through each of the five representational regions of the network's state space. As the velocity is increased the paths can undergo symmetry breaking bifurcations similar to those in the two object case. This can result in many different topologies for the paths which traverse several representational regions, and we can expect various forms of masking, ambiguity, and bias effects similar to those found in the two object case, and possibly new phenomena as well. The response of these networks to changing inputs is thus much more complicated than a simple fixed point analysis of the closed system would suggest.

5 Conclusion

We have seen that even when simple neural networks are exposed to simple model environments, complex and revealing representational structures emerge. How close the agent is to an object, what other objects it has recently observed, the velocity with which objects are passing by, and the nature of the transitions between objects all emerge quite naturally as distinct types of paths even in very simple networks. We have seen how the interplay of intrinsic agent dynamics and environment dynamics gives rise to such well known perceptual phenomena as masking and ambiguity with biasing effects. It is hoped that further study and development will allow for these tools to be applied in a wider variety of contexts, and ultimately to the study of actual neural systems in their natural environment.

References

Beer, R. [2000] "Dynamical approaches to cognitive science," *Trends in Cognitive Sciences*, **4:3**, 91-99.

Bregman, A. [1990] *Auditory Scene Analysis: The Perceptual Organization of Sound* (MIT Press, Cambridge, MA).

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

Chiel, H., & Beer, R. [1997] "The brain has a body: adaptive behavior emerges from interactions of nervous system, body, and environment," *Trends in Neuroscience*, **20**, 553-557.

Churchland, P, and Sejnowski T. [1992] *The Computational Brain*, (MIT Press, Cambridge, MA).

Clark A. [1997]. "The dynamical challenge," *Cognition*, **21:4**, 461-481.

Clark A., & Chalmers, D. [1998] "The extended mind," *Analysis*, **58:1**, 7-19.

Cohen, J. [2004] "Information & content," in *Blackwell Guide to the Philosophy of Information and Computing*, Floridi L. (ed.), 215-227. (Blackwell, New York, NY).

DeBaggis, H. F. [1952] "Dynamical systems with stable structures," in *Contributions to the Theory of nonlinear Oscillations* **2**, Lefschetz S. (ed.), 37-59, (Princeton Univ. Press, Princeton, NJ).

Hasselblatt B., & Katok A. [2002] *Handbook of Dynamical Systems* Volume 1A (Elsevier, North Holland).

Hirsch, M. W., & Smale, S. [1974] *Differential Equations, Dynamical Systems, and Linear Algebra*, First Edition (Academic Press, New York, NY).

Hodgkin, A. L. & Huxley, A. F. [1952] "A quantitative description of membrane current and its application to conduction and excitation in nerve" *Journal of Physiology* 117, 500-544.

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.

Hutchins, E. [1890] *Cognition in the Wild.* (MIT Press, Cambridge, MA).

James, W. [1890] *The Principles of Psychology*, http://psychclassics.yorku.ca/James/Principles/index.htm (accessed April 2009).

Kant, I. [1999] *The Critique of Pure Reason.* Guyer, P. & Wood, A. (trs). (Cambridge University Press, Cambridge, UK).

Kirsch, D. & Maglio P. [1994] "On distinguishing epistemic from pragmatic action," *Cognitive Science*, **18**, 513-549).

Kirsch, D. [1998] "The intelligent use of space," *Artificial Intelligence*, **73**, 31-68.

Koch, C. [2004] *The Quest for Consciousness: A Neurobiological Approach.* (Roberts and Company, Englewood, CO)

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.

Peixoto, M. M. [1959] "On structural stability," *Ann. of Math.*, **69:1**, 199-222.

Pressnitzer, D., & Hupé, J. [2006] "Temporal Dynamics of Auditory and Visual Bistability Reveal Common Principles of Perceptual Organization," *Current Biology*, **16**, 1351-1357.

Robinson, C. [1995] *Dynamical Systems: Stability, Symbolic Dynamics, and Chaos* (CRC Press, Boca Raton FL).

Rumelhart, D., McLelland, J., & the PDP Research Group. [1986] *Parallel Distributed Processing*, Volume 1. (MIT Press, Cambridge, MA).

Schecter, S. [1985] "Persistent unstable equilibria and closed orbits of a singularly perturbed system," *J. Differential Equations*, **60**, 131-141.

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 University Press, Oxford, UK).

Tuller, B., Case, P., Ding, M., & Kelso J.A.S. [1994] "The nonlinear dynamics of speech categorization," *Journal of Experimental Psychology: Human Perception and Performance*, **20**, 3-16.

van Gelder, T. & Port. R. [1995] *Mind as Motion: Explorations in the Dynamics of Cognition* (MIT Press, Cambridge, MA).

van Gelder, T. [1998] "The dynamical hypothesis in cognitive science," *Behavioral and Brain Science*, **21**, 615-665.

Warren, W. [2006] "The dynamics of perception and action," *Psychological Review*, **113**, 358-389.

Wertheimer, M. [1912] "Experimentelle studien über das sehen von bewegung," *Z. Psychol.*, **61**, 161-265.

Wiggins, S. [1990] *Introduction to Applied Nonlinear Dynamical Systems and Chaos* (Springer).

Wilson, H. [1999] *Spikes, Decisions, Actions: The Dynamical Foundations of Neuroscience* (Oxford University Press, Oxford, UK).