Active Internalism and Open Dynamical Systems[*]

Jeff Yoshimi

ABSTRACT: The question whether cognition is subserved by internal processes in the brain (internalism) or extends into the world (active externalism) has been vigorously debated in recent years. I show how internalist and externalist ideas can be pursued in a common framework, using (1) *open dynamical systems*, which allow for separate analysis of an agent's intrinsic and embodied dynamics, and (2) *supervenience functions*, which can be used to study how low-level dynamical systems give rise to higher-level dynamical structures.

The debate over active externalism (the thesis that the mind literally extends in to the world) is a bit of a draw.[1] Externalists have launched an impressive assault on tradition, citing a wide variety of empirical results (and producing philosophical arguments) that suggest that many forms of cognitive activity extend in to or are offloaded in the world.[2] Internalists have vigorously defended tradition, noting (among other things) that the brain processes information in a distinctive way, as compared with external artifacts like notebooks and cell phones (Rupert, 2004), and that the brain performs important control theoretic functions, often independently of environmental coupling (Grush, 2003).[3] Moreover, there is a long-standing tradition that assumes that the immediate causal basis of conscious experience is in the brain. The debate has raged,

---

[*] Penultimate draft (with corrections) of https://www.tandfonline.com/doi/abs/10.1080/09515089.2011.569919

[1] I refer to this debate as the "embodiment debate." I take "externalism" to be a covering term for positions that emphasize the importance of the environment in cognitive processes. "Externalism" in this sense covers active externalism, semantic externalism, embodiment, the enactive approach to perception, etc. When a more specific approach is in question, and the context does not make this clear, that position is referred to using appropriate terminology.

[2] Recent treatments, with references to earlier work, are Noë (2004), Gibbs (2006), Clark (2008), Chemero (2009), and Kiverstein and Clark (2009).

[3] For overview of the internalist response, see Adams and Aizawa (2008), and Rupert (2009). The internalist position is also treated in detail in Clark (2008).

is currently active, and is, I suspect, just past its apex. Concessions have been made on both sides. Defenders of internalist tradition almost universally acknowledge that the externalist evidence is impressive. Externalists have noted that the brain *is* in some sense distinctive; it is, as Clark says, "the core and… most active element" in the extended system (Clark, 2008, p. 139).

So, there does not seem to be much work left to do. Perhaps the right thing to do is to let the debate fade away—another section in philosophy of cognitive science anthologies. But there is work left to do. I will argue that, by combining a class of mathematical objects called *open dynamical systems* (Hotton and Yoshimi, 2010; Hotton and Yoshimi, forthcoming) with supervenience relations construed functionally (Yoshimi, forthcoming), all relevant considerations can be accommodated in a unified framework, which has advantages over partisan positions with concessions added.[4]

The resulting approach, which I refer to as "active internalism," can be thought of in terms of a sequence of steps:

Step 1: Specify a *total system* of interest, T
Step 2: Specify an *agent system* A in T
Step 3: Consider higher level-phenomena that supervene on the state space of A[5]

[4] Similar approaches exist in the literature. Clark (2008) endorses externalism for much cognitive processing but is also an internalist about consciousness (2009). Wilson (2010a) describes a "moderate externalist view", "a kind of pluralistic view of the mind vis-à-vis the debate over individualism, whereby individualistic and externalist views of cognition divide the mind between them." Wilson (2001) also provides a set of tools for distinguishing different aspects of realization, which is similar to the set of tools described here. Bechtel (forthcoming) shows how mechanistic explanation often requires simultaneously considering internal dynamics and external couplings, at multiple levels of analysis. Also note that the tools I use—what I call "open dynamical systems" and "supervenience functions"—involve formalization of concepts associated with existing areas of research (more in section 3).

[5] When I refer to a "system" or "object" I mean a real thing or collection of things, something like a spatio-temporal particular. I denote systems (and their parts) by upper case letters. When I refer to "states" and "state spaces", by contrast, I mean ways these

These steps make it possible to compare an agent's intrinsic dynamics with its open dynamics, and to study the way base dynamics induce higher level supervenient behaviors. The framework is meant to be flexible, so that the same physical system can be analyzed in multiple ways, relative to different senses of "cognition." For example, a person's *beliefs* (understood functionally) might supervene on the states of a system A_1, while the *conscious states* of the same person might supervene on the states of a system A_2, where A_1 and A_2 overlap in space, and where each is embedded in a different total system.

In what follow I make substantial use of dynamical systems theory, but I do not thereby endorse the "dynamical hypothesis" in cognitive science (van Gelder 1998). Though I am favorably disposed to that position, the approach taken here is more inclusive (connectionist and even symbolic AI approaches to cognition can be pursued in this framework).

I begin (in sections 1 and 2) by outlining the dialectical situation, canvassing the most persuasive arguments on either side of the debate. In section 3 I describe open dynamical systems and supervenience functions and show how they can be used to analyze open and closed dynamics at multiple levels of analysis. In section 4 I show how these tools can be used to pursue externalist and internalist ideas in a common

---

systems can be. More specifically I take state spaces to be a special kind of property set, and states to be members of such sets (see Yoshimi, forthcoming). I denote state spaces by upper case bold letters. For example, if T is a total system, and A is an agent system "in" T (i.e. a mereological part of T), then we can talk about a state space $\mathbf{S}_1$ for A, which is a set of ways A could be, and a separate state space $\mathbf{S}_2$ for T, which is a set of ways T could be. As we will see, multiple state spaces can apply to the same system.

framework. I also show how this framework can be used to resolve objections faced by existing positions. I conclude by considering general objections to pluralist approaches.

1. Externalist Considerations

There are now many lines of evidence in support of various forms of externalism. Rather than attempting an exhaustive review, I'll give a few representative examples.

The main evidence for the externalist position derives from cases which suggest that various types of mental processes extend beyond the brain.[6] Such cases date back at least to Merleau-Ponty (1962; first published in 1945), who criticized classical theories of perception on the basis of analysis of specific examples, such as the case of a blind man and his cane, where the proddings of the cane—and so, in some sense, the cane itself—are constitutive of the man's understanding of his environment. These cases suggest that motor activity and sensori-motor interaction are essential components of perception. In psychology, this type of argument is famously associated with Gibson (1986; first published in 1979), who claimed, on the basis of a detailed analysis of the empirical data concerning visual perception, that the visual system isn't "in the head," but is a feature of an "animal-environment system" (p. 225). The philosopher John Hagueland (1996), following Merleau-Ponty and Gibson (and borrowing an example from Herbert Simon[7]) describes an ant's walk across a sand dune, which involves not just the ant's nervous system, but also its legs and sensory apparatus, and the shape of the dune itself. It is not

[6] As Rupert (2010) says, "Proponents of the extended view often rest their case on observations about dependence" (p. 2). Also see Rupert (2004), section II.
[7] See Simon (1996; first published in 1969), p. 52.

so much that the ant knows how to get to its destination, but rather that the ant-body-dune system is such that the ant makes its way to its destination. Hagueland says that such cases involve a kind of causal "intimacy," where "the term 'intimacy' is meant to suggest more than just necessary interrelation or interdependence but a kind of *commingling* or *integralness* of mind, body, and world—that is, to undermine their very distinctness" (p. 208).

In recent years many forms of causal intimacy have been studied. For example, in cases of "offloading" cognition, a subject's performance at various tasks improves with access to external artifacts. Long multiplication using pencil and paper provides a familiar example (see Rumelhart et al., 1986)—many arithmetical problems (e.g., 4389 divided by 87) could not be completed by most subjects without pen and paper. Maglio et al. (1999) have shown that subjects can produce more words in *Scrabble* from a set of letters when they have the actual tiles before them and can manipulate them. Kirsh and Maglio (1994) famously showed that subjects perform better in the video game *Tetris* when they actively manipulate blocks than when they perform these operations in their heads. Kirsh (2010) has recently expanded on these analyses, arguing that cognitive enhancements like drawings, schematics, videos, and "marked" movements in dance choreography serve to, among other things, "provide a structure that can serve as a shareable object of thought" (p. 1).[8] These cases suggest that external artifacts are often essential components of cognitive processes.

---

[8] Also compare what Donald (1991) calls "exograms" (e.g., written texts) which serve as external stores of memory (also see Sutton, 2010).

Advocates of the "enactive" approach to perception emphasize the way perception changes—sometimes ceasing altogether—when normal environmental interactions are disturbed. For example, kittens deprived of an ability to move during development fail to develop normal depth perception. Images artificially stabilized on the retina fade from view. Subjects wearing inverting glasses are "experientially blind" until they have had a chance to develop the right kinds of sensori-motor skills. These examples and other cases (reviewed in Noë, 2004) suggest that normal perception requires bodily engagement with the world.

Others have focused less on agent / artifact combinations, than on larger structures which can themselves be viewed as cognitive systems. Hutchins (1995) has studied multi-agent systems and argued that cockpits and navigation teams (complex objects comprising artifacts and one or more human agents) "have interesting cognitive properties in their own right" (p. xiii). In such cases, he says, we should "move the boundaries of the cognitive unit of analysis out beyond the skin of the individual person and treat the navigation team [or cockpit] as a cognitive and computational system" (p. xiv). In a similar way one can consider an entire legal system as an institutional structure which produces cognitions in the form of legal judgments (Crisafi and Gallagher, 2010).

Perhaps the most famous philosophical argument for externalism is due to Clark and Chalmers (1998). The central premise of their argument is the "parity principle" according to which, for any state or process *x*, if *x* is considered cognitive when it occurs in a subject's head, *x* should also be considered cognitive when it occurs (at least partially) in the environment.[9] In the famous example of Otto and Inga, *x* is finding one's way to a destination (e.g., a museum) using stored memories. Inga finds her way to the museum using brain traces, and we have no problem considering this to be a cognitive process. By the parity principle, when Otto finds his way to the same museum with the help of a piece of paper, we should also consider the process to be cognitive. Thus, some cognitive processes extend beyond the brain and into the environment.

This argument shows that externalists are, as Clark himself says, committed to a kind of functionalism, whereby the implementing medium of a cognitive process does not matter (whether it be in the brain or in the world).[10] What maters is the way relevant information is processed. Thus, we should shed our "bio-prejudice" (Clark, 2008); we should stop assuming processes are cognitive solely because they occur in the brain and focus instead on their functional form. This has recently been referred to as "extended functionalism", "now with an even broader canvas for multiple realizability than ever before" (Kiverstein and Clark, p. 2; also see Wheeler, 2010).

Other cases could be added to this list, and the externalist position could be taxonomized and organized in different ways. See, e.g., Gibbs (2006), Menary (2006), Sutton (2006), Clark (2008), Rupert (2009), Wilson and Clark (2009), Kiverstein and Clark (2009), and Wilson (2010a). There are also additional arguments that I have not covered, for example, Wilson's meaning-making argument (Wilson, 2010b).

---

[9] This was only labeled the "parity principle" later. See, e.g. Clark (2008). Compare Adams and Aizawa (2008) on "cognitive equivalence." For critical discussion of the parity principle, see Menary (2006), Wilson and Clark (2009), and Sutton (2010). The issues raised in these papers are addressed in section 4.

[10] See Clark (2008), p. 88.

2. Internalist Arguments

I distinguish two forms of internalist argument: synchronic and diachronic.[11] Synchronic (or *static*) internalist arguments say that an individual's mental state at a time is determined by properties internal to that individual at that time. Diachronic (or *dynamic*) internalist arguments say that internal systems (in particular, nervous systems) are characterized by distinctive causal and dynamical profiles. In both cases, we will see that what counts as an "internal" system can vary from case to case.

Synchronic internalism is, in the contemporary parlance, the claim that a person's mental states locally supervene on his or her internal physical states, but *not* on wider environmental states. The standard formulation of supervenience says that a set of properties **A** supervenes on a set of properties **B** iff two objects being different with respect to the properties in **A** implies that they are different with respect to the properties in **B**. Alternatively, we can say that an object's **B**-properties at a time determine its **A**-properties at that time. So, the claim is that internal physical states, unlike environmental states, uniquely determine mental states (more on this in section 3, where *supervenience functions* are defined). [12] This was the default position in philosophy until recently, and

[11] For helpful review, history, and analysis of internalism, see Bartlett (2008).
[12] One property of supervenience is that (roughly), if **A** supervenes on **B** then **A** also supervenes on any set of properties that in some sense "includes" **B**. For example, if conscious states supervene on brain states, then conscious states also supervene on world states (if changes in conscious states entail changes in brain state, they thereby entail changes in the world those brains are part of). Thus, it is sometimes stipulated that a given base set of properties be *minimal*, i.e. the "smallest" set of properties **B** that **A** supervenes on (e.g., for the case of conscious states, this minimal set might contain states of some subsystem of the brain). This concept of a "minimal subvenient base" has not, to my knowledge, ever been made precise. Throughout when referring to "subvenient bases" I have this kind of minimal subvenient base in mind. For example, when referring

continues to be the default position in psychology and the cognitive sciences. Neuroscientists, for example, don't produce papers attempting to show that the brain is the immediate basis of consciousness (though they do try to identify *what* in the brain correlates with consciousness).[13] However, the evidence for synchronic internalism has been accumulating since antiquity, when it was first observed that mental states of agents systematically co-vary with the states of their brains, but do not covary in the same way with states of the environment.[14] The accumulation of this type of evidence has accelerated in recent years, with the explosive growth of cognitive neuroscience. Across a wide variety of domains, it is known that changes in mental state predict changes in specific regions of the brain, but do not predict environmental changes in the same way. If my mental state changed, something in my brain *must* have changed, but the external environment need not have changed (it is always possible that the change in mental state came about by direct intervention, e.g. neural stimulation).[15]

---

to the brain as the subvenient base of consciousness, I assume that there is actually some mereological part of the brain which is the minimal subvenient base of consciousness.

[13] As Bartlett says, there was no "organized defense" of internalism until recently; in fact, the position was initially formulated as a foil for critics to attack.

[14] Among Greek and Roman scholars there was a debate between those who thought the heart was the seat of the soul (e.g., Aristotle), and those who thought the brain was (e.g., Plato). The Roman physician Galen is said to have settled the issue. In his most famous demonstration, he removed the laryngeal nerves from a pig, and observed that this eliminated vocalization but no other functions. Other experiments showed that damage to specific parts of the brain had a specific impact on cognition (e.g., inducing blindness), even if the sense organs were intact (see Gross, 1988).

[15] Clark (2009) is critical of this type of evidence, wrongly I think, but I will not respond to his arguments here (though I will respond to related worries in section 4). Clark also describes a novel way of supporting a neural basis for consciousness. Citing an earlier suggestion by Chalmers, Clark notes that relatively high bandwidth communication channels within the brain (as compared with lower bandwidth channels connecting the brain to its environment) make it the "only adequate 'vehicle'" for consciousness (p. 983).

While the synchronic internalist arguments emphasize the way internal states determine mental states at a particular time, diachronic internalist arguments emphasize the fact that an agent's internal system (in particular, its central nervous system) processes information in a distinctive way, as compared with the way the body and external artifacts do.

The intuitive idea is straightforward: information processing in the brain—via a webbed array of synaptic connections—is much different from information processing in various external artifacts, e.g. the writing of notes on notepads or the storing of numbers in a cell phone. This in turn casts doubt on externalist efforts to dispense with the brain as a distinctive explanatory kind. The distinctive nature of neural processing is empirically well supported. Rupert (2004) considers a number of examples, including "recency effects," whereby humans are more likely to remember recently encoded memories than more distantly encoded memories. Of course, this is not the case with a cell phone or a pad of paper (every phone number in my cell phone is equally accessible). Similarly across a range of long-standing results of cognitive psychology, including chunking, priming, Weber's law, learning time, access time, interference, and generation effects (see Rupert 2004). These are not properties of external artifacts unless those artifacts are explicitly programmed to behave in the relevant way. This in turn suggests that, in the case of memory, there are "two different explanatory kinds, *internal memory* and *external resources used as memory aids*" (Rupert, 2004, p. 418).[16]

[16] Compare Grush's (2003) arguments concerning the control-theoretic functions of the nervous system.

Of course, nothing is special about the brain here. It is standard practice to divide the world up into distinct explanatory kinds, and to separately consider their ways of behaving. Water is a natural kind, which behaves differently from the various minerals composing soil (so that we take the boundary of a lake seriously; see Rupert, 2010, p. 30). Moreover, acknowledging that water and soil are distinctive does not preclude analysis of their interactions with other things. Indeed, it is standard scientific practice to consider the way a given system or medium (e.g., water, or a given type of mineral or soil) behaves relative to a variety of interactions.

3. Open Dynamical Systems and Supervenience Functions

In this section I describe the tools I will use to integrate the externalist and internalist considerations just canvassed. I begin with a review of dynamical systems theory, then describe *open dynamical systems*, which account for an agent's interactions with an environment. Finally, I describe *supervenience functions*, which can be used to study the way a system's lower-level dynamical behavior is related to its higher-level behaviors.

The tools I describe are related to several existing lines of cognitive research, which have a considerable history and breadth. What I call an "open dynamical system" formalizes and generalizes ideas found in Ashby (1960), Bingham (1988), Harvey et al. (1997), Beer (2000), Kelso (2001), Warren (2006), and much of the work reviewed in Harvey et al. (2005). Researchers in evolutionary robotics model the relationship between an agent's intrinsic dynamics and its dynamics when it is in an environment, and have used such models to study various cognitive phenomena, including selective attention (Slocum et al., 2000), categorical perception (Beer 2003), and communication (Marocco et al., 2003). What I call a "supervenience function," a mapping from a low-level state space to a higher-level states space, corresponds (in the case where the low-level space is a continuum and the high-level space is finite) to a kind of "coarse-graining" of the lower-level space into a finite set of regions. This idea is used in at least two areas: the area of symbolic dynamics, where continuous dynamics are associated with discrete symbol strings, and in the analysis of how finite state machines can be implemented by recurrent neural networks. In both cases the relevant ideas have been applied to problems in cognitive science.[17] For example, Izquierdo, Harvey, and Beer (2008) evolved neural networks to perform associative learning tasks in two kinds of environment and found that successful networks implemented simple state machines.[18]

3.1 Dynamical Systems

A dynamical system is a mathematical structure that describes the way a system changes over time.[19] It is essentially a rule that associates current states in a *state space*

---

[17] On symbolic dynamics in relation to cognitive science, see Dale and Spivey (2005) and Atmanspracher and beim Graben (2008). On the relationship between recurrent neural networks and state machines, see Casey (1996), which includes references to earlier work. On both topics see beim Graben, P., & Potthast (2009). These ideas are discussed in relation to the philosophical literature on supervenience in Yoshimi (forthcoming).

[18] Also see (Phattanasri, Chiel, & Beer, 2007).

[19] Dynamical systems can be formally defined (see Hotton and Yoshimi, 2010, for a discussion of the formal definition) but we will not consider the formal definition here. Also note that there is a common equivocation at this point, between the formal mathematical structure and the real system being modeled; see van Gelder (1998). The equivocation is mostly harmless and will be ignored in what follows.

(the set of all possible states for a system) with unique future states in that space. Thus, dynamical systems are deterministic—they associate initial states with unique futures—even though they can display seemingly unpredictable or *chaotic* behaviors. Many existing mathematical models in science are dynamical systems; for example, most differential equations are dynamical systems, as are most computer simulations.

One advantage of dynamical systems is that they allow us to think about a system's behavior using a visually intuitive framework. In particular, we can visualize the set of possibilities for a system in terms of its state space. The state space of a dynamical system is often an *n*-dimensional Euclidean space (or a region of such a space), which is can be thought of as the set of possible combinations of values for *n* variables. For example, the state space for a 2-node neural network is a region of $\mathbf{R}^2$ (see figure 1), the set of possible patterns of activity across the network's two nodes. For a connectionist model of brain with 100-billion neurons, the state space is a region of $\mathbf{R}^{\text{100-billion}}$, the set of possible patterns of activity across that brain's 100-billion neurons.

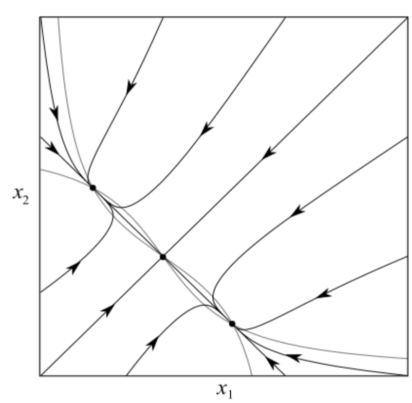

Figure 1: Phase portrait in the state space of a Hopfield network.

The state space of a dynamical system is filled with *orbits* or *paths*.[20] Each path corresponds to one way the system can evolve in time. The *phase portrait* for a dynamical system is the complete set of paths associated with it.[21] A phase portrait provides a kind of visual snapshot of the dynamics of a system, a picture of its possible behaviors. Choose any initial state, and the phase portrait shows what subsequent states will occur. Phase portraits are useful, because they are uniquely associated with dynamical systems, and they describe a system's behaviors in an intuitive way. For example, to understand the behavior of the neural network shown in figure 1, we don't need to see the equations—we can just inspect its phase portrait. Most initial states are drawn towards one of two *stable fixed points* or *attractors* for that system.

When a two-dimensional system has fixed point attractors, we can think of its phase portrait in terms of a ball rolling on a curved surface. The ball corresponds to the current state, the surface corresponds to the state space, and the path followed by the ball corresponds to an orbit. Place the ball anywhere on the surface (begin in any initial condition) and the ball will travel to the bottom of a local valley. Of course, if the dynamics were different, the phase portrait would look different. For example, a predator-prey system defined on the same state space, $\mathbf{R}^2$, has different dynamics. Instead of paths leading to the bottom of a local valley, its phase portrait contains concentric loops, which describe oscillations in predator and prey populations.

[20] The term "orbit" is standard in dynamical systems theory. The term "path" is neutral between classical dynamical systems and *open dynamical systems*, and so I will prefer that term here.

[21] I am using the phrase "phase portrait" in a somewhat non-standard way. The phrase usually refers to a visual depiction of selected orbits of a dynamical system (e.g., figure 1). I also use it to refer to a logical rather than a graphical object—namely, the full set of orbits of a dynamical system, what is sometimes called an "orbit space."

3.2 Open Dynamical Systems

A classical dynamical system is closed, in the sense that it is not exposed to outside influences.[22] On the other hand, cognitive systems (like most systems in nature), are open, in the sense that they are coupled to complex and changing environments. How can we model this kind of situation using dynamical systems theory? In (Hotton and Yoshimi, 2010; Hotton and Yoshimi, forthcoming) the issue is addressed using *open dynamical systems*.[23] The basic idea is easy to motivate if we contrast the phase portrait of a classical closed system with the phase portrait of an open system. Paths in the phase portrait of a closed system cannot cross. If they did, multiple futures would be possible from the cross points, which would violate the deterministic nature of a dynamical system. Place a ball on the curved surface corresponding to a closed attractor system, and it will go in one direction only: on a path towards the bottom of the nearest valley. Paths in an *open phase portrait*, by contrast, can cross. From a given point in an open system's state space, multiple futures are possible, depending on what happens in the environment. Imagine that the ball in the attractor example is made of steel, and that a magnet is

[22] Of course, outside influences can be incorporated, but then they are taken to be part of a larger dynamical system which is not subject to external influences (a useful strategy, as we're about to see).

[23] As noted above, others have analyzed the dynamics of agents in environments. Our contribution is to formalize and generalize these ideas using an explicit mathematical framework, which in turn provides (among other things) for a more rigorous analysis of certain dynamical concepts, e.g. hysteresis. We also describe tools for analyzing some of the complex structures that occur in an open system. These points are elaborated in Hotton and Yoshimi (2010) and Hotton and Yoshimi (forthcoming). In those papers we also differentiate open dynamical systems in our sense from open thermodynamic systems and open loop control systems.

suspended above it, influencing the course it takes. Rather than simply moving to the bottom of a local valley from a given point, the ball can now travel in multiple directions, depending on the behavior of the magnet. What results is a more complex phase portrait, a set of paths which can crisscross one another in complex ways.

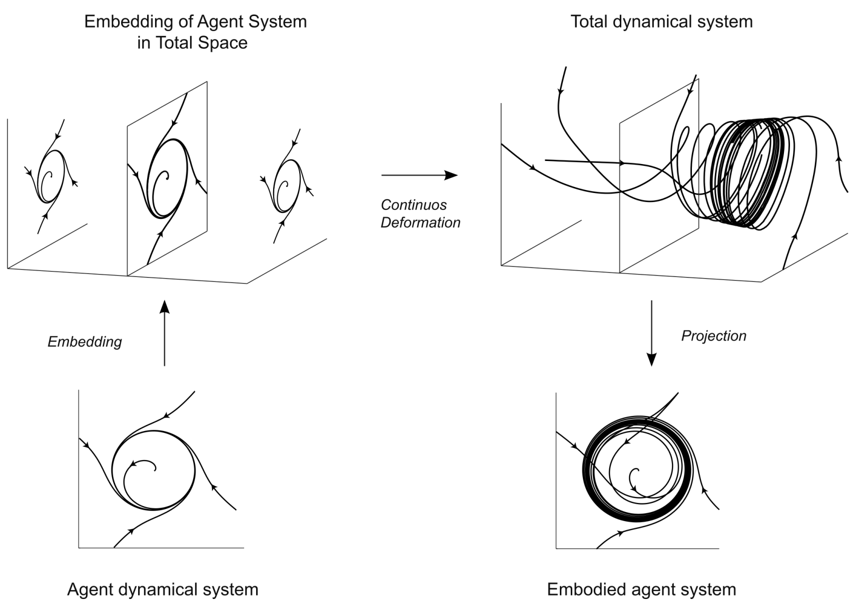

Figure 2: Schematic of an idealized open dynamical system.

Formally an open dynamical system is a compound system which separately models an agent when it is isolated from any environment, and the same agent when it is embedded in an environment. In figure 2 a simple mathematical model illustrates the main ideas. This is not a model of any actual cognitive system; it is simply meant to illustrate how the main components of an open dynamical system fit together. The four panels of the figure correspond to the main components of an open dynamical system.

We can consider these components in a clockwise movement, from the closed system on the lower left to the open system on the lower right.[24]

1) Lower left: This panel depicts an *agent dynamical system* or *agent system* defined on an *agent state space*. The agent state space corresponds to the set of possible states for some particular subsystem of a larger system (e.g. one neuron relative to the brain, one person in a group of people). The agent system models the *intrinsic dynamics* of this subsystem, the way it behaves when isolated from environmental influences. Note that the orbits in the agent system's phase portrait don't cross (compare the phase portrait for the neural network in figure 1): an agent system is a classical, closed dynamical system. In the idealized example shown in the figure, the intrinsic dynamics of the agent system contains an attracting limit cycle: begin anywhere in the agent state space and the system will end up following a loop-shaped path.

2) Upper left. This panel depicts an embedding of the agent system in a *total state space.* The total state space corresponds to the set of possible states for the larger system in which the agent is embedded (the brain as a whole, a group of people, etc.). In the idealized case shown in figure 2, the attracting limit cycle in 2 dimensions has been embedded in a 3 dimensional space. The dynamics of the agent system have not changed at this stage. The agent system is simply being placed in a higher dimensional space.[25]

---

[24] In what follows, I distinguish *open dynamical systems*, which are compound systems like the one shown in figure 2, from *open systems* or *embodied agent systems*, which correspond to one part of an open dynamical system, namely the agent system when it is embedded in an environment (see the bottom right panel in figure 2).

[25] The act of "placing" the agent system in the total space involves more formalism than this (Hotton and Yoshimi, forthcoming), but the intuitive idea should be clear. In this case, a dynamical system that consists of an infinite collection of copies of the agent system is defined on the total space. Placing the agent system in the total space then amounts to associating the agent system with one of these copies.

This in turn helps conceptualize the relationship between the agent system and a *total system*, discussed next.

3) Upper right: This panel depicts a *total dynamical system* or *total system* defined on the total state space. The total system describes the behavior of the environment (which includes the agent of interest); it models the kind of mind-body-world system Gibson and others in the externalist tradition describe. Like the agent system, the total system is a classical, closed dynamical system: its orbits don't cross. In the idealized system shown in figure 2, the total system is an attracting limit cycle in three dimensions: begin anywhere in the state space and the system will end up following a loop-shaped path. This is extremely simple, compared with the kind of complexity that actually occurs when modeling an environment together with one or more agents, but again, it illustrates the main points in a simple way. The total system's dynamics is related to the agent system dynamics by a continuous deformation of the embedded agent system. Thus far, notice, we simply have two classical dynamical systems, and a way of seeing how they are related (by an embedding and a continuous deformation).[26]

4) Lower right. This panel shows the open phase portrait of an *open system* or *embodied agent system*, which is defined on the agent state space. Each path in this phase portrait is a projection of an orbit of the total system. Each projected path thus describes one way this agent will behave, relative to a particular environmental configuration. For example, an open system could describe all possible behaviors of an animal relative to a particular animal-environment system. By comparing closed and open phase portraits in

[26] Technically there is also a third classical dynamical system, corresponding to the embedding of the agent system in the total state space.

an agent space (e.g., the lower-left and lower-right panels in figure 2), we see how the intrinsic dynamics of an agent are changed in a particular environment. [27] Open systems are my primary interest here, since they can be used to study the way internal and external forces contribute to an agent's behavior, in a context which emphasizes internal agent states.

With these tools in hand, we can study an agent's behavior (1) on its own, and (2) in the context of an environment, and we can also (3) compare its open and closed behaviors. As an example, consider the Hopfield network discussed above. Considered as a closed system, the network has an attractor dynamics whereby, from most initial states, the system settles in to one of two stable fixed points (see figure 1). We can then embed this network in an environment that involves two objects passing the network at regular intervals, altering the activations of its two neurons (Hotton and Yoshimi, 2010). The network together with the environment correspond to a total system defined on a 4-dimensional state space (the two network dimensions and two dimensions describing various configurations of objects). The open phase portrait for the network relative to this total system is shown in figure 3. Each of the paths corresponds to one possible sequence of network states, relative to a particular environmental process (the objects stimulating the network's nodes in a particular way). Compare the closed and open phase portraits shown in figures 1 and 3. The nicely delimited collection of orbits in the closed phase portrait in figure 1 has been replaced by a family of overlapping loops in figure 3, which

[27] Compare Wilson's (2010a) concept of "functional gain", a property of integrated systems whereby they display novel functions compared to the behavior of any of their components considered in isolation. Also see Wilson and Clark (2009),

must be shown in separate panels since they overlap one another.[28] Before, we had an attractor dynamics: all states tended towards one of two fixed points. Now we have a modified attractor dynamics: all states are pushed away from and *around* the fixed points. Metaphorically, the moving magnet described above is pushing the ball from valley to valley, and around the basin edges.

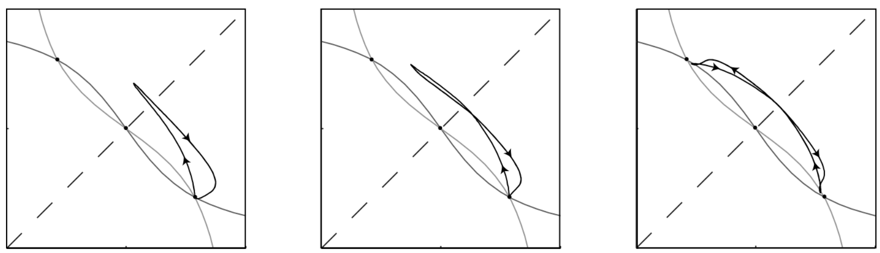

Figure 3: Paths in the open phase portrait of the open Hopfield network.

The overall behavior of the embedded network can be understood as a combination of endogenous and exogenous forces—intrinsic and environmental dynamics. The attractors of the closed system exert an influence, pulling the network towards the stable states, but external influences prevent the network from settling in to those states, so that the network is pulled in loops through the agent space.

3.3 Supervenience Functions

Having outlined a way of studying a system's closed and open dynamics, we now consider those dynamics at various levels of analysis. It is typically assumed that the lower-level states at of a system synchronically determine its higher level states. For

[28] In (Hotton and Yoshimi 2010; Hotton and Yoshimi forthcoming) techniques for dealing with this complexity are described (e.g. grouping paths into collections of paths that do roughly the same thing, and then color coding these collections).

example, the overall state of the elementary particles in my brain at time *t* fixes the state of the molecules in my brain at *t*, which in turn fixes the state of the neurons in my brain at *t*, etc. So, if we have dynamics at one level, that should determine dynamics at higher levels, relative to synchronic determination or realization relations. [29]

We can formalize these ideas using the concept of a *supervenience function* $f: \mathbf{B} \to \mathbf{A}$, which associates states in a base state space or *base space* **B** with states in a supervenient state space or *supervenience space* **A** (the class of objects which bear the states in **B** are the *subvenient base* for the supervenience function).[30] Given such a function, we can associate any base state $b \in \mathbf{B}$ with the unique supervenient state $f(b) \in \mathbf{A}$ it determines. For example, if **B** is the space of brain states, **A** is the space of human conscious states, and if human conscious states supervene on brain states, then any brain state b determines a unique human conscious state $f(b)$.[31] Supervenience functions are

[29] See (Yoshimi, forthcoming) for discussion of the relation between supervenience and dynamics, and for review of authors who distinguish diachronic causal process and synchronic determination relations. In the context of the embodiment debate, this kind of distinction has been invoked with aims similar to my own. Adams and Aizawa (2001, 2008) distinguish coupling (which is dynamic) from constitution (which is static). Clark (2009), drawing on Hurley and Block, distinguishes "instrumental dependence" (dynamic) and "constitutive dependence" (which is static; p. 965). More in section 4.

[30] I have discussed elsewhere how standard formulations of supervenience imply this functional formulation (Yoshimi, 2010). Varieties of supervenience such as strong, weak, and global (see McLaughlin and Bennett, 2005, for review) imply different functional formulations. For example, strong supervenience entails a single supervenience function prefixed by a modal parameter (so that base states determine the same supervenient states in all possible worlds), while weak supervenience entails a family of supervenience functions, one for each possible world. These differences can largely be bracketed here, though in a fuller treatment varieties of supervenience (including explicit modal parameters) could usefully enrich the overarching framework, with its emphasis on adaptability to different cases.

[31] The function need not be defined on the entire domain; there are most likely brain states that don't determine any conscious state, and similarly in other cases of supervenience. This introduces trivial complications, which I deal with elsewhere.

often many-one; many brain states can determine or realize the same conscious state, since we could have brain states $b_1, b_2, \ldots, b_n$ such that $f(b_1) = f(b_2) = \ldots = f(b_n)$. Thus we can capture the phenomenon of multiple realization in a straightforward way.

3.4 Open Dynamical Systems Relative to Supervenience Functions

Open dynamical systems and supervenience functions can be used together, to analyze the way open dynamics at one level give rise to corresponding behaviors at other levels. Consider the case of a finite supervenience space which supervenes on a continuum of base states. In such a case we can picture the supervenience function by partitioning the base space into cells, where each cell is the pre-image of a supervenient state (see figure 4). All the base states in a given cell are mapped to the same supervenient state by f (they are the "multiple realizers" of that supervenient state). Now suppose we have an open phase portrait on the base space. Such a case is shown in figure 4, where a supervenience function is superimposed on an open phase portrait (though only one path of that open phase portrait is shown). As a path of the open phase portrait unfolds, travelling from one cell to another, a corresponding path unfolds in the supervenient space.[32] In this case, a repeating loop in the base space gives rise to a repeating cycle of eight supervenient states.

[32] Note that, while the supervenient state set in this example is discrete, this need not be the case; it could be a continuum.

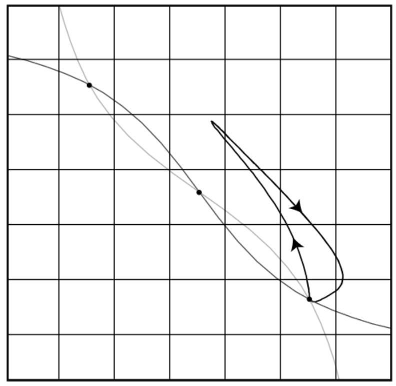

Figure 4: One path of an open phase portrait, relative to a supervenience function.

These tools are flexible, insofar as they can be used to analyze multiple regions of an object or region of space. For example, figure 5 schematically depicts a set of objects, A_1 - A_4, some of which are parts of one another, some of which overlap (i.e. share a part). Various open systems / supervenience analyses could be pursued here. We could, for example, treat T_1 as a total system, and then consider any of A_1 - A_4 as agent systems relative to this total system. Or, we could treat A_1 and A_2 together as a total system, T_2. Each of these agent systems and total systems is associated with a state space, and in some cases these spaces are associated with a supervenience function. For example, following Clark, and using figure 5, we can suppose that Otto's conscious states supervene on his brain states (states of A_1, let us say), while his beliefs supervene on combined states of Otto and his notebook (states of T_2).[33] So, in this case we have two supervenience functions: one from the state space of A_1 to the space of conscious states

[33] As Clark says, "inscriptions in the notebook figure as part of the physical supervenience base for certain standing beliefs of the agent" (2010, p. 966).

for Otto, and another from the state space of T_2 to the space of beliefs for Otto. There are other supervenience relations here as well. For example, visual experiences are thought to supervene on states of a functionally-defined region of the brain, while neural firing rates supervene on states of individual neurons. So, within this one physical system, we have multiple, overlapping subvenient bases (a brain-artifact combination, a brain, a functional region of the brain, and individual neurons), relative to multiple supervenience functions.[34]

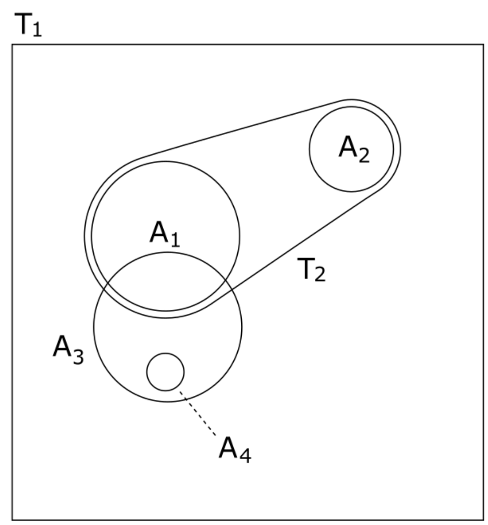

Figure 5: A region of space that can be analyzed in multiple ways.

4. Active Internalism

[34] Matters are even more complex than this. Having specified a base space of interest, it is sometimes the case that a hierarchy of spaces supervenes on that base space. For example, physical states of the brain determine chemical states, which determine biological states, which sometimes determine psychological states. This situation has an interesting analysis in terms of the partitions described above, whereby as one moves up the hierarchy of supervenience functions, increasingly coarse-grained partitions are induced on the lowest level base space.

Recall the steps described at the beginning of the paper:

Step 1: Specify a *total system* of interest, T
Step 2: Specify an *agent system* A in T
Step 3: Consider higher level-phenomena that supervene on the state space of A

We have seen in some detail how these steps work: we have seen what a total system and agent system are, and how supervenience functions can be used to analyze the way they give rise to higher-level behaviors. In this section I apply these ideas to the embodiment debate, showing how the seemingly opposed ideas presented in sections 1 and 2 can be integrated in this framework. I also use the approach to address issues and objections that have emerged in the literature.

4.1 Integration of Externalist and Internalist Considerations

Much of the evidence for externalism reviewed in section 1 is based on cases where an extended system (a system extending beyond the brain) is associated with cognitive properties. In this framework, such systems can be thought of as total systems of an open dynamical system (step 1). We thus have total systems corresponding to: a blind man together with his cane; an ant together with a sand dune (or any other Gibsonian "animal-environment system"); a person with his or her cell phone; a complete game of *Scrabble* including players and tiles; a ship with its sextants and navigators; a legal system including persons, documents, and institutions; Otto with his notebook; a person with pen and paper solving a math problem; etc. There are interesting differences between these cases. In some cases, the total system is a kind of transient agent / artifact assembly; in other cases the total system is more stable and persistent (cf. Wilson, 2010a, who refers to "systems of various levels of durability and robustness over time and circumstance"; also see Wilson & Clark, 2009).

The diachronic internalist considerations discussed in section 2 emphasize the distinctive way certain systems processes information: nervous systems, for example, clearly process information in a distinctive way, as compared with the way cell phones and notebooks do. To capture this, we model the intrinsic dynamics of a particular system by treating it as the agent system of an open dynamical system (step 2). This preserves the notion that particular systems or media (e.g., the nervous system) behave in distinctive ways, which are essential to our understanding of them. For example, in the Otto and Inga case, we can treat Otto as an agent system relative to a total system encompassing him and his notebook. In this way we can study the intrinsic dynamics of his (impaired) neural memory system, and then separately study the way those dynamics change (the "functional gain" that occurs) when Otto is coupled to his notebook.

According to the synchronic internalist arguments discussed in section 2, certain types of mental states—e.g. conscious states—locally supervene on internal physical states, but not on wider environmental states. We have seen how supervenience can be applied in this framework. The state space of any agent system or total system can be the base space of a supervenience function (step 3). For example, we can take the state space of the brain (an agent or total system, depending on the context) to be the base space of a supervenience function which associates brain states with conscious states.

By combining these steps we can study a wide range of internalist and externalist phenomena in a unified framework. For example, suppose human conscious states supervene on brain states, and consider the case of a navigator working on a ship. To understand the conscious processes that occur for this navigator, we first consider the ship as a kind of cognitive system in its own right (Hutchins, 1995), by treating it as the total system of an open dynamical system. We also consider the intrinsic dynamics of the navigator's brain, and in this way acknowledge that as a trained human navigator, she and her neural circuitry process information in a distinctive way (diachronic internalism). Finally, we take seriously the notion that conscious states are locally determined by brain states (synchronic internalism) and thus view the dynamics of her conscious experience as being determined by the open dynamics of her brain, relative to a supervenience function. In this way we can study the way the phenomenology of ship navigation is based on the open dynamics of brain activity, by way of a local supervenience function.

A second example emphasizes the role of constructs familiar from the classical internalist tradition, in a framework that also recognizes the importance of embodiment. Insect locomotion depends in crucial ways on the insect's body and environment.[35] This complicates central controller schemes, according to which animal behavior is completely governed by the operation of a central mechanism. Acknowledging this, we can begin our analysis with a total system that models an insect's brain together with its body and environment. But we can also recognize that understanding insect locomotion involves understanding what happens *inside the insect*. One mechanism widely believed to be essential is a central pattern generator, a neural circuit which sends repeating pulses of activity to the insect's limbs. These circuits are understood in part by detaching them

---

[35] Recall that this was Hagueland's example in discussing "causal intimacy." For review see Chiel and Beer (1997).

and studying them isolation. [36] When the circuits are decoupled in this way, they produce "fictive motor patterns":

> By far the most compelling argument that a piece of the nervous system is intrinsically able to generate a rhythmic motor pattern is to remove it from the animal and place it in a dish filled with physiological saline. Under these conditions there are no sensory pathways remaining, and no timing information available from the environment. Today, many preparations have been shown to generate what are called fictive motor patterns, motor patterns that would drive muscle movement if the muscles were present (Marder and Bucher, 2001).

Having described the fictive motor patterns produced by an isolated insect—that is, having described (this aspect of) the intrinsic dynamics of the insect *qua* agent system—we can ask what changes when the insect is embedded in an environment. That is, we can ask how the dynamics of the isolated insect are affected by its interaction with a real environment. In fact, this is already of interest to workers in the field. Much of the review cited above addresses the question: "How closely do fictive motor patterns resemble those generated in the intact animal during movement?" (p. 987). I speculate that in this case we would observe a set of closed loops (corresponding to the repeating pattern produced by the central pattern generator) transformed into a more complex set of overlapping loops reflecting environmental feedback and interaction.

In different cases, different aspects of the analysis are more or less important. Some cases emphasize agent systems with important intrinsic dynamics (e.g. central pattern generators); others emphasize a total system (e.g., ship navigation); others emphasize supervenient phenomena (e.g. social facts, the wide content of mental states). Not all the steps described above need be followed, nor must they be followed in the

---

[36] Compare the analysis of dissociated cortical cultures (which allow isolated brain tissue to be studied), or for that matter just about any analysis of a biological system *in vitro*.

order suggested. In the insect case, for example, we have an agent system with important intrinsic dynamics, but there are no obviously interesting supervenient phenomena (though some might later be discovered). Conversely, we might begin with a supervenient phenomenon, but determine its subvenient base *not* to be interesting as an agent or total system its own right. For example, the value of a dollar bill supervenes on a gerrymandered spatio-temporal region encompassing relevant value-conferring institutions, but this region does not seem to be independently interesting as a dynamical system.

4.2 Response to Subsequent Debate

We are now in a good position to see how a variety of arguments and issues that have emerged in the embodiment literature can be addressed in this integrative framework.

4.2.1 The parity principle

Recall that, according to the parity principle, if a state or process is considered cognitive when it occurs in an agent's head, it should also be considered cognitive when it occurs outside of an agent's head. The principle has been the subject of considerable debate (see note 9), which has focused attention on the question of how cognitive processes should be demarcated. In judging for parity, should the emphasis be on fine grained causal profiles (such that Otto and his notebook constitute distinct explanatory kinds) or on more abstract functional properties (such that Otto and his notebook should be treated as a single extended cognitive system)? Does parity imply similarity (such that a truly *cognitive* notebook should retrieve memories in a manner similar to the way a brain does), or not? Moreover, what makes some process cognitive to begin with, such that questions about parity can get started? Proposals that have been made in the context of the embodiment debate include Rupert's (2004) proposal that "something is cognitive if and only if it is part of a persisting, integrated cognitive system," Adams and Aizawa's (2008) claim that cognitive processing involves causal operations on "non-derived content", and Wheeler's (2010) defense of a traditional functionalist definition of mental states.[37]

According to the pluralist perspective advocated here, multiple principles of demarcation should be simultaneously considered (cf. figure 5 above). According to some of these principles, Otto and his notebook together form a cognitive system. According to other principles (e.g., principles emphasizing fine grained functional profiles) only Otto by himself, or some part of him, is cognitive. In this way we can study multiple aspects of cognition as well as relations between them.

4.2.2 The coupling constitution fallacy

A prominent objection to externalists is that they commit a "coupling-constitution" fallacy (Adams and Aizawa 2001), which involves an invalid inference

---

[37] Also see Clark's (2008) discussion of "coarse grained functional poise," and Bartlett's (2008) discussion of "core realizers."

from constitution from coupling. For example, just because the bimetallic strip in a thermostat is coupled to the air in a room, does not make the room part of (constituted by) the thermostat. To claim that the thermostat's coupling to the room makes the room part of the thermostat is to commit the coupling-constitution fallacy. The blind-man's cane is coupled to the blind-man, but this does not imply that his conscious perception is *constituted* by the cane (that is, it does not follow from the man's reliance on his cane that the cane is literally part of his perceptual process). Of course, the distinction between coupling and constitution is central to this framework—coupling corresponds to the relationship between an agent and its environment (which together comprise a total system), and constitution relations can be described by supervenience functions. So, this framework is well-suited to studying embodied cognition without committing the coupling-constitution fallacy.

Externalists have responded to the charge that they commit this fallacy in several ways. One response is to deny that it captures the most important patterns of inference at work in externalist arguments. As several authors have pointed out (Menary 2006, Wilson 2010a), one need not begin with some cognitive object *x*, then ask whether items *become* cognitive by their coupling to *x*. Rather, multiple systems together create "an integratively coupled system that is a causal entity in its own right" (Wilson, 2010a). The present framework accommodates this perspective: an integratively coupled system can be considered a total system in its own right. However, this is compatible with recognizing the distinction between coupling and constitution, and the danger of conflating them. For example, an integratively coupled system could itself be coupled to a wider environment, where that (second form of) coupling could be mistaken for constitution. So, we can acknowledge that in some cases a system is integratively coupled, while also maintaining that the coupling-constitution fallacy is important to bear in mind and avoid.

4.2.3 Enactive approaches

The enactive approach to perception (reviewed in section 1) motivates a "dynamic entanglement" argument against synchronic internalism, according to which the way conscious processes unfolds depends so much on the brain's embodiment and embedding, that there may be no such thing as a "minimal internal neural correlate."

> [We] conjecture that consciousness depends crucially on the manner in which brain dynamics are embedded in the somatic and environmental context of the animal's life, and therefore that there may be no such thing as a minimal internal neural correlate whose intrinsic properties are sufficient to produce conscious experience (Thompson and Varela 2001, quoted in Cosmelli and Thomson (forthcoming)).

In support of this, Cosmelli and Thomson provide a detailed account of the various ways in which brains are embodied and embedded. They also give a mathematical argument: "neuronal and extraneuronal state variables are so densely coupled as to be nonseparable."

In response to the second argument, note (1) that there is no mathematical reason—no reason having to do with coupling of state variables—why orbits in the state space $\mathbf{S}_T$ of a total system can't be projected to a subspace of $\mathbf{S}_T$ (an example of a such a projection is shown in figure 2). Moreover, (2) any subspace of a state space can, at least in principle, be the base space of a supervenience function. So coupling of state variables provides no formal reason to reject the possibility of local supervenience.

Moreover, the more general argument in Cosmelli and Thomson does not provide evidence against "minimal neural correlates", and the steps described above provide a clear way of seeing why. [38] We can (1) acknowledge "dynamical entangling" between brain, body, and world, by beginning with a total system which corresponds to a brain-body-world system. We can then (2) project the orbits of this total system to the state space **B** of the brain, in this way generating a collection of overlapping paths, which reflect the impact of dense coupling. Finally, (3), we can consider a supervenience function from **B** to the space **C** of conscious states, and use this function to map the tangled, overlapping paths of **B**, to potentially even more tangled and overlapping paths in **C**. Thus, the radically open dynamics of the brain can give rise to equally if not more radically open dynamics of consciousness, *by way of a local supervenience function*. That is, we can acknowledge the way "consciousness depends crucially on… the somatic and environmental context of the animal's life" but deny that this provides evidence against minimal neural correlates; just the opposite: we use the minimal neural correlates to understand this dependence.

4.3 Limitations of the Framework

Though I have emphasized the generality of this framework—its ability to accommodate multiple competing perspectives—I do not believe it is completely general; I don't think it can easily accommodate all forms of cognitive phenomena. It is well

[38] Clark, 2009, counters this argument in a similar way; see pp. 970-987. Also see the discussion of "Searle's objection" in Noë, 2004, p. 219ff.

suited to the study of systems that are relatively well delimited in space-time. A brain is an example, as is a neuron, or the human body. But some systems are in various ways gerrymandered and dispersed. The subvenient base of a dollar bill's value is a widely dispersed set of value-conferring institutions and social structures. There is no obvious agent system in such a case. Or again, Clark (2008, ch. 6) refers to "soft assembled, temporary medleys of information processing resources" (p. 116).[39] Such assemblies cannot easily be detached and treated as separate units of analysis. To analyze such cases new tools may be needed (perhaps we already have one of those tools with the concept of "soft assembly"). On the other hand, I believe such tools could be used in the context of this framework. For example, the soft-assembly of resources could be understood as occurring within the agent system of an open dynamical system.

5. Conclusion

I have described a set of tools, and a framework for applying them to the analysis of cognitive phenomena. Using these tools, I have tried to show how seemingly opposed ideas in the embodiment debate can be pursued in a common framework. The idea was not to be neutral, in the sense of being open to various positions, nor was the purpose to restate the embodiment debate in more precise terms. Rather the purpose was to show how multiple positions in this debate are simultaneously true; to show how both sides of the debate have it (mostly) right. So, I have described a form of pluralism, something akin to political centrism or dialectical synthesis, which sees the truth in multiple

[39] On "soft assembly" and related concepts, see Bingham (1988).

positions and tries to integrate them all.[40] Such approaches face certain general difficulties, which I end by addressing.

One objection to such positions is that, in trying to make everyone happy, they end up backpedaling and equivocating (contrast the tough, decisive actions of *realpolitik*). I'm not sure anyone would make such an objection in print, but I've encountered the attitude, and it's worth highlighting the obvious response: the world is complicated, and we need appropriately complex tools to study it. If that means current positions have to be refined ("backpedaling"), so be it. If a term has multiple meanings, those meanings need to be distinguished and separately analyzed. Uncompromising debate serves a purpose in clarifying conceptual issues, but in the long run the hard work of studying cognition—in all its overlapping forms and guises—has to happen.

Another problem faced by centrist positions is that they are not clearly distinct from partisan positions, suitably amended. As we saw in the introduction, participants on both sides of the debate make important concessions to their opponents, and externalists in particular make many of the points I've made here (cf. note 4). However, the position I have described is weighted more towards internalism than any modified externalism that I am aware of.[41] In particular, some of the clearest targets of externalism—central controller schemes, internal representations, within-the-head realization bases for consciousness and other cognitive phenomena—are vigorously endorsed here. The phrase

---

[40] Pluralist approaches to cognitive science have been a subject of considerable recent discussion. See, e.g., Spivey and Anderson (2008), Dale et al. (2009), and McLelland et al. (2010).

[41] Of course, the tools I've described are neutral with respect to the philosophical positions. In particular they could be used by someone with strictly internalist or externalist leanings.

"active internalism" is thus apt: important things happen *in the head*. The brain is a repository of representations and computational structures, and is the subvenient base for consciousness. However, these internally determined structures are also actively conditioned by their embodiment in the world, so that we can study the open dynamics of internal representations, or the open dynamics of consciousness.[42]

---

[42] I am grateful to Bill Bechtel, Rick Dale, Scott Hotton, Georg Theiner, an anonymous referee, and audiences at two conferences where earlier versions of this paper were presented (*Embodied, Embedded, Enactive and Extended Cognition*, at the University of Central Florida, and a meeting of the *Central Valley Philosophy Association*, at College of the Sequoias).

References

Aizawa. K. (2007). Understanding the Embodiment of Perception, *Journal of Philosophy*.

Adams, F. and Aizawa, K. (2001). The Bounds of Cognition, *Philosophical Psychology*. 14:1, 43-64.

Adams, F. and Aizawa, K. (2008). *The Bounds of Cognition*. Blackwell.

Ashby, W. R. (1960). *Design for a Brain*. Chapman and Hall London.

Atmanspracher, H. and beim Graben, P. (2008). Contextual Emergence of Mental States from Neurodynamics. *arXiv*, q-bio/0512034v1 [q-bio.NC].

Bartlett, G. (2008). Whither Internalism? How Internalists Should Respond To The Extended Mind Hypothesis, *Metaphilosophy*. 39:2.

Bechtel, W. (forthcoming). The Downs and Ups of Mechanistic Research: Circadian Rhythm Research as an Exemplar, *Erkenntnis*.

Beer, R. (2000). Dynamical Approaches to Cognitive Science, *Trends in Cognitive Sciences*, 4:3, 91-99.

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

beim Graben, P., & Potthast, R. (2009). Inverse problems in dynamic cognitive modeling. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 19, 015103.

Bingham, G. (1988). Task-Specific Devices and the Perceptual Bottleneck, *Human Movement Science* 7, 225-264.

Block, N. (2005). Review of Alva Noë, *Action in Perception*, *Journal of Philosophy*. 259-272.

Casey, M. (1996). The dynamics of discrete-time computation, with application to recurrent neural networks and finite state machine extraction. *Neural computation*, 8(6), 1135-1178.

Chemero, A. (2009). *Radical Embodied Cognitive Science*. MIT Press.

Chiel, H., and 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.

Clark, A. and Chalmers, D. (1988). The Extended Mind, *Analysis* 53, 7-19.

Clark. A. (2007). Curing Cognitive Hiccups, *Journal of Philosophy*, 104, 163-192.

Clark, A. (2008). *Supersizing the Mind*. Oxford: Oxford University Press.

Clark, A. (2009). Spreading the Joy? Why the Machinery of Consciousness is (Probably) Still in the Head. *Mind*, 118, 963-993.

Cosmelli, D. and Thomson, E. (forthcoming). Embodiment or Envatment? Reflections on the Bodily Basis of Consciousness. In Stewart, J., Olivier G, and di Paolo, E. (Eds.) *Enaction: Towards a New Paradigm for Cognitive Science*, MIT Press.

Crisafi, A. and Gallagher, S. (2010). Hegel and the extended mind, *AI & Society*, 25, 123-129.

Dale, R., Dietrich, E., & Chemero, A. (2009). Explanatory Pluralism in Cognitive Science. Cognitive Science, 33(5), 739-742.

Dale, R. and Spivey, M. (2005). From apples and oranges to symbolic dynamics: a framework for conciliating notions of cognitive representation, *Journal of Experimental and Theoretical Artificial Intelligence*, 17, 317-342.

Donald, M. (1991). *Origins of the modern mind: Three stages in the evolution of culture and cognition*. Harvard Univ Pr.

Gibbs, R. (2006). *Embodiment and Cognitive Science*. Cambridge: Cambridge University Press.

Gibson, J. (1986). *The Ecological Approach to Visual Perception*. New York: Taylor and Francis.

Gross, C. (1988). Galen and the Squealing Pig, *The Neuroscientist*, 4, 216-221.

Grush R. (2003). In Defense of Some "Cartesian" Assumptions Concerning the Brain and its Operation, *Biology and Philosophy*, 18, 53-93.

Harvey I., Husbands P., Cliff D., Thompson A. and Jakobi, N. (1997). Evolutionary Robotics: the Sussex Approach, *Robotics and Autonomous Systems*, 20, 205-224.

Harvey, I., Di Paolo, E, Wood, R., Quinn, R. Tuci. E. (2005). Evolutionary Robotics: A New Scientific Tool for Studying Cognition, *Artificial Life*, 11, 79-98.

Haugeland, J. (1996). Mind embodied and embedded. In *Having Thought: Essays in the Metaphysics of Mind*. Harvard University Press.

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

Hotton, S., & Yoshimi, J. (2010). The Dynamics of Embodied Cognition. *International Journal of Bifurcation and Chaos*, 20(4), 1-30.

Hotton, S., & Yoshimi, J. (forthcoming). Extending Dynamical Systems Theory to Model Embodied Cognition. *Cognitive Science*.

Izquierdo, E., Harvey, I., & Beer, R. D. (2008). Associative learning on a continuum in evolved dynamical neural networks. *Adaptive Behavior*, 16(6), 361-384.

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

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

Kirsh, D. & Maglio, P. (1994). On distinguishing epistemic from pragmatic action. *Cognitive Science*, 18, 513-49.

Kirsh, D. (2010) Thinking with external representations. *AI & Society*.

Kiverstein, J. and Clark, A. (2009). 'Introduction: Mind Embodied, Embedded, Enacted: One Church or Many?' *Topoi*, 28, 1-7.

Maglio, P., Matlock, T., Raphaely, D., Chernicky, B., and Kirsh, D. (1999). Interactive skill in Scrabble. In *Proceedings of Twenty-First annual conference of the cognitive science society*. Lawrence Erlbaum.

Marocco, D., Cangelosi, A., & Nolfi, S. (2003). The emergence of communication in evolutionary robots. *Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 361(1811), 2397-2421.

Marder, E. and Bucher, D. (2001). Central Pattern Generators and the control of rhythmic movements, *Current Biology*, 11, 986-996.

McClelland, J. L., Botvinick, M. M., Noelle, D. C., Plaut, D. C., Rogers, T. T., Seidenberg, M. S., & Smith, L. B. (n.d.). Letting Structure Emerge: Connectionist and Dynamical Systems Approaches to Cognition.

McLaughlin, B. and Bennett, K. (2005) Supervenience, in Edward N. Zalta (Ed.) *The Stanford Encyclopedia of Philosophy.*

Menary, R. (2006). Attacking the bounds of cognition, *Philosophical Psychology*, 19, 329-344.

Merleau Ponty, M. (1962). *The Phenomenology of Perception*. C. Smith, (Tr.) Routledge.

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

Phattanasri, P., Chiel, H. J., & Beer, R. D. (2007). The dynamics of associative learning in evolved model circuits. *Adaptive Behavior*, 15(4), 377-396.

Rumelhart, D.E., McClelland, J.L., Smolensky, P., Hinton, G.E. (1986). Schemata and sequential thought processes in PDP models. In D.E. Rumelhart, J.L. McClelland, and the PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of Cognition*, Vol. II., *Psychological and biological models* (7–58). Cambridge, MA: MIT Press.

Rupert, R. (2004). Challenges to the Hypothesis of Extended Cognition, *Journal of Philosophy*. 101: 389-428.

Rupert, R. (2009). *Cognitive Systems and the Extended Mind*. Oxford University Press.

Rupert, R. D. (2010). Extended cognition and the priority of cognitive systems. *Cognitive Systems Research*.

Simon, H. A. (1996). *The sciences of the artificial*. The MIT Press.

Slocum, A. C., Downey, D. C., & Beer, R. D. (2000). Further experiments in the evolution of minimally cognitive behavior: From perceiving affordances to selective attention. *From animals to animats*, 6, 430-439.

Sober, E. (1999). Physicalism from a probabilistic point of view, *Philosophical Studies* 95 (1-2):135-74.

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

Sutton, E. (2006). Introduction: Memory, Embodied Cognition, and the Extended Mind, *Philosophical Psychology*, 19, 281-289.

Sutton, E. (2010). Exograms and Interdisciplinarity: history, the extended mind, and the civilizing process. In R. Menary (Ed.), *The Extended Mind*, MIT Press.

Thompson, E. and Varela, F. J. (2001) Radical embodiment: neural dynamics and consciousness, *Trends in Cognitive Sciences*, 5, 418-425.

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

Wheeler (2010). In defense of extended functionalism. In R. Menary (Ed.), *The Extended Mind*, MIT Press.

Wilson, R. and Clark, A. (2009), How to Situate Cognition: Letting Nature Take its Course" in Aydede, M. and Robbins P. (Eds.), *The Cambridge Handbook of Situated Cognition*, 55-77.

Wilson, R. (2001), Two Views of Realization, *Philosophical Studies*, 104, pp.1-30.

Wilson, R. (2010a). Extended Vision, in N. Gangopadhyay, M. Madary, and F. Spicer (Eds.), *Perception, Action and Consciousness*. New York: Oxford University Press.

Wilson, R. (2010b). Meaning Making and the Mind of the Externalist, in R. Menary (Ed.), *The Extended Mind*. MIT Press.

Warren, W. (2006). The Dynamics of Perception and Action. *Psychological Review*, 113, 358-389.
Yoshimi, J. (forthcoming). Supervenience, Dynamical Systems Theory, and Non-Reductive Physicalism. *British Journal for the Philosophy of Science*.