Supervenience, Dynamical Systems Theory, and Non-Reductive Physicalism[*]

Jeff Yoshimi

ABSTRACT: It is often claimed (1) that levels of nature are related by supervenience, and (2) that processes occurring at particular levels of nature should be studied using dynamical systems theory. However, there has been little consideration of how these claims are related. To address the issue, I show how supervenience relations give rise to "supervenience functions", and use these functions to show how dynamical systems at different levels are related to one another. I then use this analysis to describe a graded approach to non-reductive physicalism, and to critically assess Davidson's arguments for psychological anomaly. I also show how this approach can inform empirical research in cognitive science.

A classical and ongoing topic in philosophy of science and mind concerns the relationship between "levels" of nature (Oppenheim and Putnam [1958]; Nagel [1961]; Davidson [1970]; Fodor [1974]; Hellman and Thompson [1977]; Kim [1984]; Horgan [1993a]; Bickle [1998]; Shagrir [1998]; Heil [2003]; Wilson [2005]; Melynk [2006]; Craver [2007]; Dardis [2008]). It is generally agreed that physical systems aggregate into increasingly complex structures, existing at different levels of organization. For example, organisms have a fundamental physical structure, as well as a chemical, biological and (in some cases) a psychological structure. Since the 1970s, relations between these levels have often been characterized in terms of supervenience relations, which are supposed to enforce inter-level dependency while allowing the sciences which study different levels to be autonomous. Another prominent topic in recent philosophy of science (in particular, philosophy of cognitive science) has been dynamical systems theory, which studies the way systems change their state over time in a mathematically rigorous and visually intuitive way (Van Gelder [1995]; Clark [1998]; Van Gelder and Port [1995]; Bechtel and Abrahamson [2006]). Dynamical models are typically framed at a single level of the

---

[*] Penultimate draft of https://www.journals.uchicago.edu/doi/abs/10.1093/bjps/axr019

kind of organizational hierarchy described above (e.g. dynamical models of an atom, protein, neuron, or psychological process). This naturally raises the question of how dynamical processes at different levels of nature relate to one another relative to the supervenience relationships that are said to connect those levels.

In what follows the relationship between supervenience and dynamical systems theory is systematically developed. By combining the two concepts we see what a dynamical system that applies to an object at one level of organization implies about that object considered at a higher or lower level of organization. For example, if we have a dynamical system on the space of brain states for an organism, and we assume that the psychological states of that organism supervene on its brain states, what does this imply about its psychology? Do the neural dynamics induce a psychological dynamics relative to the supervenience relation? Conversely, if we have a psychological dynamics for such an organism, does this constrain its neural dynamics?

In section 1 I give a philosophical analysis of the concepts of a state and a state-space, which are key to the analysis of supervenience in relation to dynamical systems theory. In section 2 I discuss the relationship between supervenience as typically defined, and "supervenience functions". I also consider multiple realization in relation to these functions. In section 3 the relationship between dynamical systems theory and supervenience is systematically developed. I show what happens to high level and low level dynamical systems relative to a supervenience function. In the remainder of the paper I put the framework developed in sections 1-3 to work. In section 4 I describe the general picture of inter-level relations motivated by this work. In particular, I describe a graded approach to non-reductive physicalism, based on quantifiable *degrees of coupling* between sciences. In section 5 I use the propositions introduced in sections 2 and 3 to argue, contra Davidson, that mental-physical supervenience conjoined with a system of physical laws implies a system of psychological laws. In section 6 I show how these ideas can be used to facilitate empirical work in cognitive science, focusing on “symbolic dynamics”, which associate continuous neural dynamics with discrete representational dynamics.

Sets (usually, sets of properties) will be denoted by upper case bold letters ‘**A**’, ‘**B**’, ‘**S**’, etc. Upper case italic letters ‘*A* ’, ‘*B* ’ ,‘*X* ’, etc. will be used as second-order variables ranging over properties in those sets. Lower case italic letters ‘*x*’, ‘*y*’, etc. will be used as first-order variables (with a few exceptions listed below) ranging over individuals which have those properties. Upper-case letters ‘A’, ‘B’, etc. will be used as constants denoting specific properties. The italic letter ‘*f* ’, will be reserved for functions and ‘*t*’ will be reserved for times.

1. States and State Spaces

Since most of the results below rely on the special characteristics of states and state spaces, and since states and state spaces have received little philosophical scrutiny (despite the philosophical use they are put to; see e.g. Churchland [1995]; Gardenförs [2000]), I here give a brief analysis.

First, I consider states to be properties in the straightforward sense that they can be instantiated by numerically distinct things. Two different apples can be red, just as two different apples can have the same mass, temperature, or velocity (where mass, temperature, and velocity are standard examples of states in science). I also assume that states are instantiated by objects at times, to capture the idea that objects can change state over time. Thus, our domain consists of objects at times.[i]

Second, when states are instantiated, they must be uniquely instantiated. An object can only have a single mass, temperature, or velocity at any given time. Were it not for this uniqueness property of states we could not describe state changes using deterministic laws, for the state of an object at a time would itself be indeterministic. To capture this, we think of properties as states when they exist in "state sets":

DEFINITION: **S** is a *state set* iff $\forall x\ \forall X_{\in \mathbf{S}}\ \forall Y_{\in \mathbf{S}}\ ((Xx \ \&\ Yx) \rightarrow X = Y)$

(where *x* ranges over objects at times). For example, if **S** is the set of temperatures, and an object at a time has two temperatures, then those are just the same temperature. That is, objects at times can only have a single temperature. Similarly with velocities, masses, and combinations thereof (e.g. only a single velocity / mass pair ever applies to an object at a time).[ii]

Third, if a set of properties is not already a state set, it can be associated with a state set, assuming conjunction and negation are allowed as property-formation operators (this is important below, for it means that anywhere supervenience obtains a supervenience function exists). This is done by associating a set of properties with the set of "maximal properties" constructible from that set (where, roughly, each property in the original set, or its negation, is conjoined into one long conjunctive property; see Kim [1984]). For example, **E*** = {happy, tall} is not a state set, for a person can be happy and tall at the same time, but the associated set of maximal properties, **E** = {happy & tall, ~happy & tall, tall & ~happy}, is a state set, because only one such conjunction applies to a person at a time (if any does).[iii] Below, when discussing mental states, it is assumed we are either dealing with maximal mental properties in this sense, or with sets of mental properties which otherwise satisfy the definition of a state set. For example, a "visual experience" is something like a maximal distribution of qualia in the visual field: only one such field can apply to a person at a time (by contrast, individual qualia are not states, since multiple qualia can apply to a person at a time). More generally, we can think of mental states as maximal mental properties, which capture *everything* relevant to a person's mental life at a time: not just a person's visual or cognitive state, but *all* of the sensory, perceptual, and other mental properties that apply to that person at that time (the complete visual field, background music, a sense of one's body, tacit beliefs, etc).[iv]

Fourth, state sets are often state-*spaces*, where the word "space" indicates that the set in question has some additional structure, in particular, some non-trivial topological structure, and generally additional structure as well. How much additional structure can meaningfully be ascribed to a state space varies by case. For example, it is plausible to assume that the set of mental states is a metric space, so that we can say some pairs of mental states are "closer" to one another than others. It is not as obvious whether it is a vector space (so that we could meaningfully speak of "adding" mental states and multiplying them by scalars). In the remainder of the paper, I assume most state sets have at least some non-trivial topological structure, so that we can speak of "supervenient spaces" and "base spaces" without further qualification.

Fifth, strictly speaking, state spaces are sets of points (mathematical objects) which symbolically represent states *qua* properties (masses, temperatures, brain states, etc.), but to avoid cumbersome rhetoric I will usually suppress this distinction, as is standard in applications of the concept of a state space. Thus, in referring to "states" and "points" in a state space, I sometimes mean the mathematical objects in the state space and sometimes mean the actual states-qua-properties they correspond to.[v]

2. The Supervenience Function and its Inverse

Supervenience is generally taken to be relation between sets of properties, **A** and **B**, where, if things are the same in terms of their **B** properties, they are the same in terms of their **A** properties.[vi] The properties in **B** are called "base" properties; the properties in **A** are called "supervenient" properties. I also refer to these as "low level" and "high level" properties, respectively.

In order to treat supervenience as a functional relation between state spaces, we first define a "determination" or "realization" relation, which is here taken to be a simple generalization linking states:

DEFINITION: F *determines* or *realizes* G iff $\forall x(\mathrm{F}x \rightarrow \mathrm{G}x)$

(where 'F' and 'G' denote states). For example, if F corresponds to a particular configuration of particles in a region of space, and G corresponds to being 20 degrees Celsius, and if F determines G, then any time the particles in a region of space enter state F, that region of space has a temperature of 20 degrees.[vii,viii]

Supervenience relations can be associated with "supervenience functions", whereby the base state of a thing uniquely determines its supervenient state:

> PROPOSITION 1: If **A** and **B** are state sets, and **A** supervenes on **B**, then there exists a function f: **B′** → **A** (where **B′** is a subset of **B**), defined by the rule $f(X) = Y$ iff X determines Y. [ix]

Given that any set of properties can be associated with a state set by construction of maximal properties, proposition 1 implies that any supervenience relation can be associated with a supervenience function f.[x] It can be shown, moreover, that f is generally an onto function (given the dependence clause associated with extant forms of supervenience; see note vi).

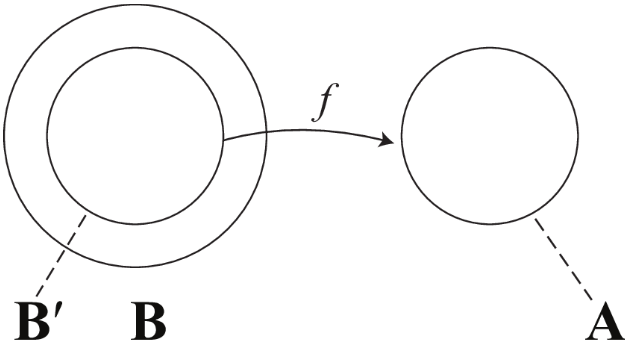

Figure 1: Schematic of a supervenience function.

To get a visual sense for how supervenience functions work, consider figure 1. If an object at a time in the domain (not shown) instantiates some state X in **B′**, then f tells us what state $Y = f(X)$ in **A** it must also instantiate. For example, if **A** contains possible human mental states, and **B** contains possible human brain states, and if it is the case that mental states supervene on brain states, then proposition 1 says that there exists a supervenience function f from **B′** onto **A**. This function tells us, for each brain state X in **B′**, that any person in brain state X at a time will also be in mental state $f(X)$ at that time. Note that there can be brain states in **B** but not **B′**. These are brain states that don't determine any mental state at all (e.g. brain states corresponding to coma or dreamless sleep). We can imagine that the sequence of brain states that occur over a person's life is a winding curve in **B**, which weaves in and out of **B′** as that person falls into and emerges out of dreamless sleep (or is knocked out, becomes comatose, etc.). So long as this neural "life curve" is in **B′**, it determines a corresponding sequence of mental states in **A,** via the supervenience function.

In other cases, f might associate physical states of art objects with their aesthetic valuations, chemical states of cells with biological states, physical states of worlds with ethical states, etc. In each case the supervenience function just represents a new way (one well suited to examining the relation between supervenience and dynamics) of thinking about the relevant supervenience relation.

Existing varieties of supervenience (e.g., strong local, weak local, strong global, weak global) can be handled in this framework, with interesting results, but these differences make little difference to the discussion which follows.[xi]

We can think about "multiple realization" in terms of the inverse of the supervenience function. Any onto function $f:\mathbf{X} \rightarrow \mathbf{Y}$ can be used to define another function which associates elements of **Y** with their preimages under f, that is, a function $f^{-1}: \mathbf{Y} \rightarrow \text{Pow}(\mathbf{X})$ such that for any $Y \in \mathbf{Y}$, $f^{-1}(Y) = \{X \in \mathbf{X} : f(X) = Y\}$. Though it is not a true inverse, I shall, following common practice in mathematics, refer to this as the "inverse" of the supervenience function (in this sense, square roots are an inverse of the squaring function: they associate numbers with collections of numbers; e.g., $\sqrt{4} = \{-2, 2\}$). The inverse of the supervenience function associates each supervenient state in **A** with those base-states in **B′** which realize it (see figure 2).

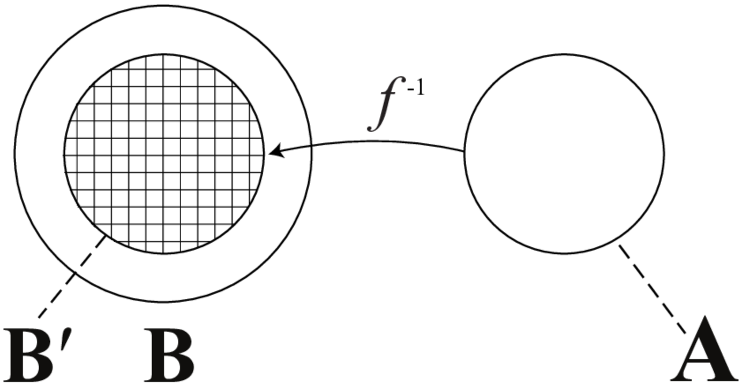

Figure 2: The inverse of a supervenience function. States in **A** are mapped to "realization classes" of **B**, which are shown as squares in a grid.

The inverse of the supervenience function induces an equivalence relation on **B′**, "realizes the same supervenience state." The members of the partition corresponding to this equivalence relation will be referred to as "realization classes." Each square in the grid in figure 2 corresponds to one realization class, all of whose elements map to the same supervenient state in **A**. Note that f^{-1} is a bijection (a 1-1 and onto function) from **A** onto the set of realization classes it induces in **B′**.

Some further points about realization and realization classes should be noted, in order to prevent misunderstandings, illustrate further directions in which a "state space" approach to inter-level relations can be taken, and facilitate the work below.

First, the realization relation has itself become a focal topic in recent years (see note vii). The concept of realization offered here is minimal—a core concept shared by most other accounts.[xii] For example, Shoemaker ([2007]), who describes a relatively sophisticated account of realization, begins with the assumption that "the realizer of a property instantiation should be metaphysically sufficient for the occurrence of that property instantiation" (p. 6). I only assume this minimal concept, though I add a mathematical framework for thinking about *collections* of realization relations (more in section 4).

Second, cases of multiple realization vary in "scope" or "strength" (Kim [1989]; Horgan [1993b]; Bickle [1998]; Shagrir [1998]). One can consider the realization of mental states in arbitrary physical systems, in mammalian brains, in human brains, in *my* brain, or even in my brain *at this time*. The transition from wider to narrower scoped cases corresponds to a transition from intuitively "larger", more heterogeneous realization classes to smaller, more coherent realization classes. In the broadest case, all possible physical realizations of mental states (human, Martian, etc.) are combined, and it's not clear that we have a coherent base "space" at all, i.e. a state set with non-trivial topological structure. In other cases, which restrict attention to specific structure types, state space representations are not only possible, but extant. For example, the case of adult human brain states has been dealt with by imaging researchers who project data from multiple human subjects to a "normalized" human brain space (e.g. the "Talairach atlas"), so that human mental states have realization classes inside of this space. In what follows I usually have the narrower-scoped cases in mind, insofar as I focus on how the tools I describe can be used to study relations between existing scientific disciplines, in particular psychology and neuroscience.[xiii]

Finally, note that sets of states (e.g. realization classes) can be interpreted in several ways: (1) some sets of states correspond to *kinds*; (2) other sets of states correspond to what I call "*bounding regions*". The two cases are easily confused, because they can coincide (a set of states can be both a kind and a bounding region), and because in both cases we can think of the relevant set of states as itself being a "state" relative to a higher order state set. It is important that we keep these interpretations distinct, because in some cases below it seems I have (1) in mind, but actually intend (2).

When a set of states corresponds to a *kind*, those states are not “wildly disjunctive”, but rather have something in common; that is, they are *similar* in some way (usually other conditions are also added, e.g., that the states be in the extension of a predicate that appears in a law of the relevant science).[xiv] In such cases, we can think of the disjunction of the states in the set as itself constituting a higher order state. For example, consider a set of human brain states, each of which involves oscillations in the 8-12 Hertz frequency range. These brain states have something in common, and there are laws that describe changes in such “Alpha waves”, and so we can think of these states as comprising a kind, a higher order state—the “Alpha wave state” (relative to a higher order state set which also includes the Beta wave state, the Gamma Wave state, etc.). In such cases it makes sense to say “the brain is *in* its alpha state”, even though the brain is actually in some more specific state which manifests the alpha wave. By contrast, a collection of states with nothing in common do not constitute a higher order state.

By a “bounding region” I mean a set of states which is such that an object at a time *must be* in one of those states.

> DEFINITION: A bounding region is a set of states **G** such that a system at a time must be in one of the states in **G**.

In most cases, an entire state space is a trivial bounding region; for example, any physical system must have some temperature, thus the set of temperatures is a bounding region for any physical system. However, what will be of interest below are cases where bounding regions constrain the behavior of a system in interesting ways. For example, we may know that given the physical structure of a system and its environment at a time, that at that time it must be in one of 15 of its 10,000 possible states—those 15 states thus comprise a non-trivial bounding region for the system at that time. Note, moreover, that the states in a bounding region need not have anything in common with one another, and thus need not be kinds. Where bounding regions are discussed below I do not assume that they must also be kinds.

3. Dynamical Systems Theory and Supervenience

In this section I consider what supervenient dynamical systems imply about base level behavior, and what base dynamical systems imply about supervenient level behavior.[xv] In the second case, we will see that even if a base system fails to induce a supervenient system, it still induces a "powerset system", whose degree of "constraint" can vary. At the end of the section these results are extended to the case of an "open dynamical system" (a dynamical system embedded in an environment).

A dynamical system is a rule which describes how a system evolves over time. Dynamical systems are pervasive in science insofar as iterated functions and most differential equations are dynamical systems. Formally, a dynamical system is defined as follows:

> DEFINITION: A map $\phi : \mathbf{S} \times \mathbf{T} \rightarrow \mathbf{S}$ is a dynamical system iff: (1) There is a time $t_0 \in \mathbf{T}$ such that for all states $X \in \mathbf{S}$, $\phi(X, t_0) = X$. (2) For all states $X \in \mathbf{S}$, and times $t_1, t_2 \in \mathbf{T}$, $\phi(X, t_1 + t_2) = \phi(\phi(X, t_1), t_2)$.[xvi]

Given an initial condition $X \in \mathbf{S}$ for a dynamical system at t_0, the map ϕ says what unique state that system will be in at all times $t_i \in \mathbf{T}$. Dynamical systems are, by definition, deterministic—initial conditions always have unique futures—though they can manifest complex and unpredictable behavior (e.g., “chaotic” dynamics).

A dynamical system $\phi : \mathbf{S} \times \mathbf{T} \rightarrow \mathbf{S}$ determines a set of “orbits” or “paths”, which are time-evolutions in **S** consistent with ϕ.[xvii] Each path of a dynamical system can be thought of as one possible way that system could evolve in time, relative to the dynamical system. A complete collection of paths for a dynamical system is called a “phase portrait.” Note that paths in the phase portrait of a dynamical system cannot cross, for if they did, multiple futures would follow from the cross-point. However, we will see that the images of a set of paths relative to a supervenience function can cross, as can the paths of what I call an “open dynamical system” below. The concept of a phase portrait is useful, for it captures everything essential about a dynamical system in a conceptually (and for state-spaces with less than 3 dimensions, visually) intuitive way. Thus, I will sometimes reason about dynamical systems by reasoning about their phase portraits.

We can separately consider the case of a supervenient system’s relation to base level behavior, and a base system’s relation to supervenient behavior. The first case is easier, insofar as a supervenient dynamical system ϕ^s can immediately be associated with a dynamical system that constrains base level behavior, using the inverse of the supervenience function, f^{-1}, which, as we saw, is a bijection onto a set of realization classes. We thus get a dynamical system ϕ^r on a set of realization classes *qua* bounding regions (but in many if not most cases, *not* kinds). This dynamical system provides a deterministic way of describing constraints on the base level behavior of a system (at any time *t*, the base system must be in the realization class specified by ϕ^r at *t*).[xviii] We will see that the notion of a dynamical system on bounding regions is important below.

However, this direction of dynamical system induction is not my main interest here, for in such cases there is little reason to separately consider base and supervenient dynamics, given that they are isomorphic copies of one another.

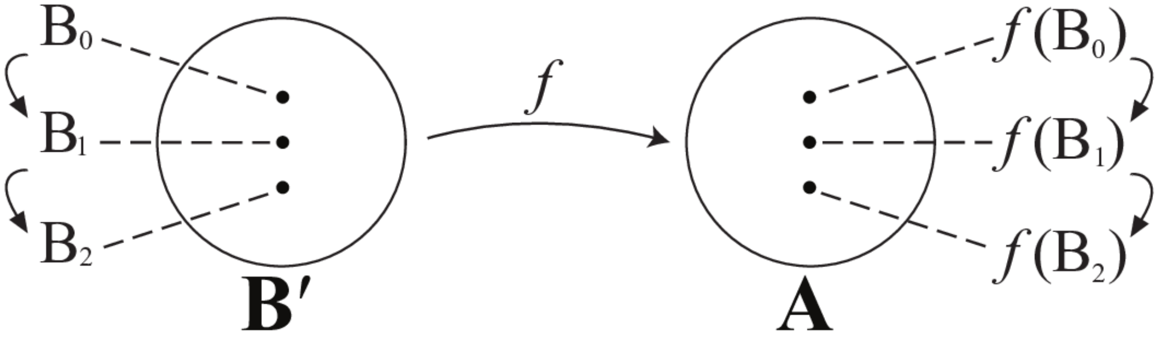

Figure 3: The supervenient image of a path of a base dynamical system.

We will focus on the more difficult and interesting case where we have a base dynamical system and want to know what it implies at the supervenient level. For simplicity, we also focus on base dynamical systems which have been restricted to **B′**. At a first pass the analysis seems simple. Since base paths are just time-ordered sequences of base states, they can easily be associated with their supervenient images under *f* (see figure 3). So, the phase portrait of a dynamical system on a base space **B′** trivially yields a supervenient analogue under the supervenience function. For example, given a neural dynamical system for a person at a time, and assuming human mental states supervene on neural states, we can say just how that person's mental states will change over time. However, does this give us a supervenient dynamical system? Can mental states themselves be associated with unique futures? Can we write down rules which allow us to predict how mental states will evolve given only knowledge of what initial mental state a person is in?

The answer is: only when a “compatibility condition” holds:

> PROPOSITION 2: A base dynamical system, ϕ^b on $\mathbf{B'}$ will induce a supervenient dynamical system ϕ^s on $\mathbf{A}$ if and only if the following “compatibility condition” holds: $\forall X, Y \in \mathbf{B'}$, if $f(X) = f(Y)$ then $\forall t\, f(\phi^b(X, t)) = f(\phi^b (Y , t))$.

(The proof is straightforward so it has been omitted). Roughly speaking, the compatibility condition says that if two states begin in the same realization class, they evolve to states in the same realization classes at all future times. This ensures unique futures for initial conditions in the supervenient state set.

Dynamical systems that are compatible in this sense are said to be “semi-conjugate.” If *f* is bijective, they are also “conjugate.” Since supervenience functions are typically many one, if a base system were to meet the compatibility condition and induce a supervenient system, the base and supervenient systems would likely be semi-conjugate rather than conjugate. In such a case the topological characteristics of the base paths could potentially be different from the supervenient paths they determine. For example, a repeating loop in the base system might get mapped to a single point in the supervenient system (e.g. an oscillating brain state might could get mapped to a fixed emotional state, without violating the compatibility condition). However, the compatibility condition is rarely if ever met in practice (i.e., in cases where real inter-level relations are in question), so it will not be emphasized here.[xix]

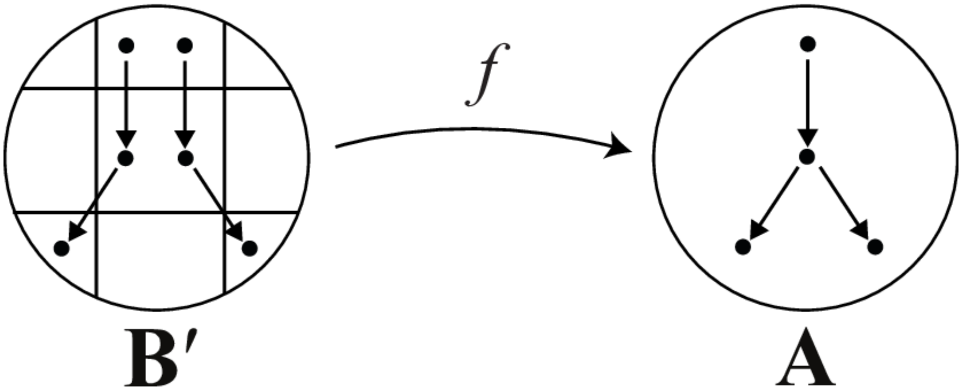

Figure 4: A case where a base dynamical system does not induce a supervenient dynamical system. **B′** is partitioned into nine realization classes. Two paths in **B′**, and their images under *f*, are shown.

Let us now consider the case, which clearly does occur in practice, where the compatibility condition *fails* to hold, so that base dynamics do *not* induce supervenient dynamics. If the compatibility condition does not apply, then two states in the same realization class can evolve to distinct realization classes at future times (see figure 4). The top two base states in **B′** in figure 4 are in the same realization class and hence map to the same supervenient state in **A**. But they are embedded in paths which lead to states in different realization classes, which therefore map to different supervenient states. The result, at the supervenient level, is that the same supervenient state can evolve to multiple future states. What is observed at the supervenient level can thus appear to be indeterministic—the top state in **A** in figure 4 will sometimes evolve one way, sometimes another–even though it is determined by a deterministic system.

In such a case, however, the *range* of supervenient states a system may visit from an initial state can still be constrained by the base dynamics. In fact, any base dynamical system ϕ^b on **B′** will induce a dynamical system ϕ^p on the *powerset* of **A**:

> PROPOSITION 3: Let **A** and **B** be state sets, and assume **A** supervenes on **B**. Let ϕ^b: $\mathbf{B'} \times \mathbf{T} \rightarrow \mathbf{B'}$ be a dynamical system on **B′**, $\mathbf{Y} \in \text{Pow}(\mathbf{A})$, and set $\phi^p : (\mathbf{Y}, t) := f(\phi^b (f^{-1}(\mathbf{Y}), t))$. Then $\phi^p : \text{Pow}(\mathbf{A}) \times \mathbf{T} \rightarrow \text{Pow}(\mathbf{A})$ is a dynamical system.[xx]

Intuitively, this means you can take an arbitrary bounding region **Y**, and use the base dynamics and the supervenience function to determine unique bounding regions at all future times *t*. The resulting dynamical system is a "powerset dynamical system." The states of a powerset dynamical system are bounding regions. The states these bounding regions contain will be referred to as "first-order states." Bounding regions are, as noted above, constraints on a system's behavior, such that a system in a bounding region **P** must be in one of the states that **P** contains. For example, if a system is in bounding region $\mathbf{P} = \{B_1, B_2\}$ at a time, it must either be in B_1 or B_2 at that time. A powerset dynamical system ϕ^p: $\text{Pow}(\mathbf{A}) \times \mathbf{T} \rightarrow \text{Pow}(\mathbf{A})$ determines unique bounding regions for all future times. This allows that from a given first-order state multiple first-order states may follow, even though bounding regions are uniquely determined.

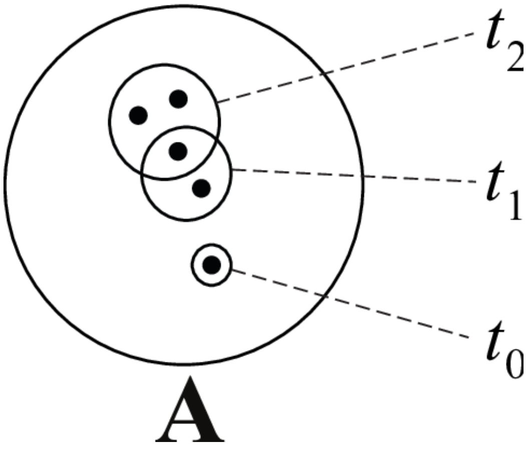

Figure 5: A simple powerset dynamical system.

For example, in figure 5, a path of a powerset dynamical system ϕ^p is shown. ϕ^p associates an initial state at t_0 (a singleton bounding region, with just one member) with a bounding region containing two first-order states at time 1, and a region containing three first-order states at time 2 (note that the bounding region at time 1 overlaps the bounding region at time 2). This says that if the system begins in the first-order state corresponding to t_0 it must be one of the two first order states determined by the powerset dynamics at time 1, and in one of the three first-order states determined by the powerset dynamics at time 2. So, a powerset dynamical system provides a deterministic way of computing what the bounding regions are for a system over time, which says given that these are your possibilities now, these will be your possibilities at future time *t*.

Although powerset dynamical system are deterministic, they vary in the degree to which they constrain first-order behavior. Proposition 3 shows that any bounding region determines a unique bounding region for all future times. However, the number of first-order states these regions contain will vary. A first-order state may evolve to every first order state at a future time *t*, or to only a single first order state at *t*. If we place the powerset system in each of its possible initial (singleton) bounding regions, and see where ϕ^P takes each of these states at time *t*, we can compute how many first order states the resulting bounding regions contain, on average, at time *t*. If we also average over future times, we can compute the "degree of constraint" of a powerset system ϕ^P:

> DEFINITION: The *degree of constraint* of a powerset dynamical system ϕ^P is the mean number of first-order states contained in its bounding regions at future times, assuming it begins in a singleton bounding region. [xxi]

This is a value between 1 and *n* (where *n* is the number of first-order states of the system the powerset system is defined on). At one extreme, every initial state can visit just 1 state at all future times, in which case we have *maximal constraint* (and in effect, the powerset system is equivalent to a first-order system). At the other extreme, every initial state can visit every state of the system at all future times, in which case we have *minimal constraint*. That is, all bounding regions at all future times just are **B′**.

We have thus far considered dynamical systems as traditionally defined, which are closed systems. However, the mind and brain, which are emphasized below, are "open systems", insofar as they engage in complex causal interactions with their environments (see, e.g., Clark and Chalmers [1988]; Noë [2004]). The formal definition of an open dynamical system is complex (Hotton and Yoshimi [2010]), but for our purposes it suffices to consider the difference between the phase-portrait of a traditional dynamical system and the phase portrait of an open dynamical system. As noted above, the paths of a dynamical system cannot cross (multiple futures would then follow from the cross-point). The paths of an open dynamical system—the members of an "open phase portrait"—by contrast, can cross (for example, see figure 7 below, where five overlapping paths of an open dynamical systems are shown in separate frames). The basic idea is that from the cross-points of paths in an open system, multiple futures are possible, depending on the state of the environment.

It is not hard to establish a generalization of proposition 3 which shows that any base dynamical system, open or closed, induces a powerset dynamical system at the supervenient level.

> PROPOSITION 4: If **A** and **B** are state sets, and **A** supervenes on **B**, then any dynamical system (open or closed) on **B** will induce a powerset dynamical system on **A**.

To see this informally, recall first, that both closed and open dynamical systems can be associated with a collection of paths on a state space (a phase portrait), with the difference that the paths of an open phase portrait can cross while the paths of a closed phase portrait cannot. Either way, a dynamical system on **B** is associated with a collection of paths in **B**, which can be restricted to **B′** and projected to **A** using *f*, and in either case what results is a set of paths in **A** which can cross. Since any collection of paths in a state space will induce a powerset dynamical system on that state space,[xxii] these projected paths induce a powerset dynamical system on **A**.

4. Inter-Level Relations and Non-Reductive Physicalism

I now turn to applications of the framework developed above. In this section I describe the general picture of inter-level relations motivated by that framework, by way of comparison with Jaegwon's Kim's work.

First, like Kim I believe that the dynamics of mental processes "piggyback" on the dynamics of physical processes, assuming mental-physical supervenience. Consider the diagrams associated with his well-known causal exclusion arguments (see, e.g., Kim [2000], [2003]), which conclude that the apparent causal powers of supervenient systems actually derive from those of lower level base processes. Figure 3 above can be seen as an instance of one of these diagrams, with the difference that I describe the supervenience relation using a function. This seems like a fairly trivial embellishment. However, a single low-level process (which is the case Kim emphasizes, for legitimate reasons) is, from my perspective, a degenerately simple case, whose simplicity masks important features of inter-level relations. My emphasis is on the complex *collections* of paths (the phase portraits) that correspond to dynamical systems, and what happens to them when they are projected between levels. When one takes this expanded purview, a whole logic of inter-level relations emerges, which has not to my knowledge been elaborated in the literature. It has not been pointed out that in some cases base dynamics generally do, and in other cases they do not, induce supervenient dynamics, or that base dynamics induce a dynamics on bounding regions at the supervenient level (and these are just the most elementary propositions of the overall logic I have in mind). Indeed, the concept of a dynamical system is hardly present in Kim or related discussions at all. But dynamical systems are routinely postulated at all levels of science, so that such an analysis is pressing. Pursuing such analyses requires moving beyond the simple two step causal sequences depicted in causal exclusion diagrams, which are useful for, e.g., the analysis of mental causation, but are insufficient for understanding inter-level relations in their full complexity.

Second, like Kim, I am doubtful about the prospects for non-reductive physicalism. In particular, I agree with Kim ([1989]) that low-level processing details manifest themselves at the supervenient level, so that reptiles, humans, and Martians probably have different psychologies (which calls into question the non-reductivist's hopes for a general psychology autonomous from implementation level concerns).[xxiii] However, where Kim (like most who discuss the issue) describes a binary choice between reductive and non-reductive approaches, I emphasize *degrees of coupling* between pairs of sciences. To assess this degree, structures in a state space associated with one science can be projected to a state space associated with another science, using a supervenience function, and the degree to which the projected structures are changed can be computed. For example, I conjecture that the physical realization classes induced by economics are more widely dispersed (see note xiii) in their base space than are the neural realization classes induced by psychology. I also conjecture that economic dynamics induced by physical dynamics are less constrained than the psychological dynamics induced by neural dynamics. So, my overall vision is one whereby discussions of the autonomy or non-autonomy of special sciences are replaced by analyses of the (measurable) degree to which structures are changed when projected between levels using supervenience functions.

5. Davidson on Psychological Anomalism

In this section I critically re-assess Davidson's ([1970], [1980a], [1980b]) argument for the anomalism of the mental, using the tools developed above. Since Davidson's argument is often taken to support non-reductive physicalism, this will allow me to further elaborate my "graded" approach to inter-theoretic relations. I argue that, in the particular case of neuroscience and psychology, there is substantially more inter-theoretic coupling than Davidson (and non-reductivists who draw inspiration from him) suggest.

Davidson assumes (1) that there are deterministic laws at the physical and neural levels. This is most evident in his discussion of "Art", an artificial replica of a person whose inner workings, including his brain, are completely understood, so that we can predict and explain all his behaviors relative to possible stimuli ([1980b], p. 245). This clearly fits the definition of an open dynamical system, which says what future states a system will be in given its current state and the state of its environment. Davidson also believes (2) that a supervenience relation connects the physical and the psychological levels ([1970], p. 214; [1980b], p. 253); indeed, Davidson is famous for being one of the first to apply the concept of supervenience in philosophy of mind. Davidson goes on to claim (wrongly, I think) that (1) and (2) are compatible with two other claims: the claim (3) that there are no psychophysical laws, that is, laws connecting physical events with psychological events, and (4) the claim that there are no deterministic laws at the mental level. (4) is his famous statement of the "anomalism" of the mental ([1970], p. 208). I take the results above to show, *contra* Davidson, that (1) and (2) are not in fact compatible with (3) and (4).

Davidson's (2) and (3) are in immediate conflict with one another, insofar as supervenience implies supervenience conditionals (proposition 1), whereby, in the case of strong supervenience, necessarily, for any brain state X and for any person with brain state X, that person will have mental state $f(X)$. That sounds like a strict psychophysical law, contra (3). This line of argument has been elaborated by Kim and Latham and has been responded to by Davidson himself, so I will not take it up further here.[xxiv]

I am more interested in a tension between (1), (2), and (4). (1) implies that the brain can be described by an open dynamical system on brain states, and (2) together with proposition 1 implies a supervenience function connecting brain states and psychological states. This, together with proposition 4, implies the existence of a powerset dynamical system on the space of psychological states, that is, a dynamical system which says, given that a subject begins in a particular mental state, what mental states that subject is constrained to be in at all future times (i.e. which bounding region applies to that agent at all future times). This is in conflict with (4), which says, in Davidson's words, that "there are no strict deterministic laws on the basis of which mental events can be predicted and explained" ([1970], p. 208) or again, "there are good arguments against the view that thought, desire, and voluntary action can be brought under deterministic laws, as physical phenomena can" ([1980a], p. 230). However, a powerset dynamical system is a set of deterministic laws at the psychological level: given that a person begins in a particular psychological state, the powerset system says exactly which psychological states are possible for that person at all future times.

Of course, the question is one of degree, for recall that powerset dynamical systems vary in their degree of constraint. If the powerset system induced by the open dynamics of brains in various environments typically have minimal constraint, then this is a hollow victory against Davidson (a minimally constrained system is such that from each of its initial states all of its possible states are possible at all future times).

However, the psychological constraint induced by open neural dynamics is not, in general, minimal. The point is intuitively obvious: of all mental states possible for me in principle, only some are available at any given time. For example, given the structure of my brain and environment, I can't have an experience of juggling bowling pins or dancing the tango before a live audience right now, even though I could have such experiences with sufficient preparation.[xxv] The point is further supported by analysis of simple model systems, e.g. the one shown in figure 6 below. In that case, a base dynamical system and a supervenience function (implicit in the realization classes) induce a powerset system on the set of supervenient states. That powerset system constrains the supervenient dynamics in a non-trivial way. If the system begins in state Q, then, given the base dynamics and the supervenience function, at all future times the supervenient system is constrained to be in states Q, X, or Y. It *cannot* be in states A, B, P, or R. Moreover, once it arrives in state X it must stay there, and similarly for state Y. So we have a powerset system at the supervenient level, with non-minimal constraint. This type of case is (in my experience modeling these situations) common; in fact, it takes deliberate effort to build model systems of this type that are minimally constrained at the supervenient level.

Davidson could respond at this point by allowing for some individual psychological regularities but denying that such regularities form a "*system* of laws which would permit the precise prediction and explanation of particular mental phenomena" (Child [1994], pp. 77-78, emphasis mine). However, with respect to prediction, this is exactly what powerset dynamics provide. Powerset dynamics correspond to systems of relations between bounding regions, which allow us to predict, to precisely defined tolerances (corresponding to degrees of constraint), how mental phenomena unfold.

Davidson could further respond as follows. He could concede that powerset dynamics with non-minimal constraint obtain and allow for prediction but deny that the resulting constraints correspond to psychologically significant regularities, i.e. psychological laws which could feature in scientific explanations. Perhaps induced powerset dynamics just give us unsystematic collections of associations, which don't reflect anything behaviorally interesting.

However, there is a vast and accumulating body of empirical evidence which suggests that (in my terms) powerset dynamics are associated with genuine psychological laws. To take just one example, specific classes of stimuli are interpreted by the visual system in predictable ways, e.g. producing specific types of illusion. The statistical generalizations described in such cases are taken to be laws which explain why subjects have particular visual experience in response to particular stimuli. The time course and statistical properties of these processes can be modeled using deterministic computational models (e.g. neural network models) together with a specification of which model states correspond to which perceptual states (i.e., a supervenience function). So, we have (1) a psychologically significant regularity—a psychological law—which explains why, when a subject is exposed to a particular stimulus, a particular visual experience or illusion occurs, and (2) a derivation of this law from neural dynamics and a supervenience function. I believe that such cases occur in every domain of human cognition, so that there are a wide variety of psychological laws that can be associated with induced powerset dynamics. However, this is just a sketch of a response to the objection concerning explanation, and further development is needed.[xxvi], [xxvii]

One thing that might be said in Davidson's defense here is that he operates with a different conception of psychology than the one I've assumed. Anomalism (4) derives from Davidson's conception of psychology, which is informed by his work in experimental decision theory (e.g., [1970], p. 220). On that conception, the question of what state an agent is in *at a given time* is itself difficult if not impossible to resolve. This is because Davidson focuses on mental state attributions, our ability to determine what state an agent is in, given the behavioral evidence (cf. Child's discussion of "interpretationism"). But I do not consider the *attribution* of mental states at all (so that there can be a question about whether a given attribution is correct or not) but simply posit a space of possible mental states for an agent and assume that an agent at a time is in one such state. Moreover, the kinds of mental states Davidson has in mind are propositional attitudes like beliefs and desires. By contrast I have focused on various mental states which capture *everything* about a subject's psychology at a time (e.g. complete visual fields or maximal psychological properties). Thus, the tension I describe may be an artifact of our different conceptions of psychology. As Davidson himself says "to the extent that psychology does not make essential use of the concepts I have described, the considerations that follow do not apply to it" ([1980b], p. 246).

However, clearly at some level a conflict has occurred, at least in spirit. The whole point of Davidson's argument is to decouple the way psychology proceeds from the way physical systems are studied, that is, to underwrite theoretical autonomy of psychology and the physical sciences, *in spite of* his overarching naturalism. But the conclusions drawn above motivate explicit interaction between neuroscience and psychology, *non*-autonomy of the two disciplines, whereby neuroscientists and psychologists work together to understand how the detailed dynamical processes that unfold in an embodied brain induce specific constraints on human psychological processes.

Ultimately, I think this framework can accommodate both what is plausible and what is problematic about Davidson's arguments. On the one hand, it is plausible to suppose, consistently with Davidson, that the constraints imposed by an induced powerset system are sufficiently non-minimal for there to be significant latitude in the way mental processes unfold over time, which could in turn account for the indeterminacy of mental state attributions. On the other hand, to the extent that induced powerset constraints are associated with genuine psychological regularities, Davidson's intuitions about the anomaly of the mental are wrong. There are remarkable systematicities in the structure of mental life, which constrain how cognition can unfold, and these constraints derive from the systematicity of the brain and its physical environment.[xxviii]

6. Symbolic Dynamics

In this section I show how the tools developed above can be applied to empirical research, focusing on "symbolic dynamics", a mathematical technique which has recently been applied to the analysis of cognitive processes.[xxix] The basic idea is to partition the state space **D** of a dynamical system into a set of regions, each of which is associated with a symbol in a set of symbols **S**. Then, as the dynamics are "run"—as an orbit in the system's phase portrait travels through the various regions—a corresponding sequence of symbols occurs (see figure 6, where **D** is a subset of $\mathbf{R}^2$, and **S** = {A, B, X, Y, P, Q, R}). This is clearly an instance of the ideas above, if we take **D** to be a base set of states, **S** to be a supervenient set of states, and the supervenience function f: $\mathbf{D} \rightarrow \mathbf{S}$ is defined as $f(P \in \mathbf{D})$ = the symbol in **S** associated with the region P lies in.

The concept of symbolic dynamics originated in mathematics, where a theorem that is hard to prove about a dynamical system is sometimes more easily proven by focusing on an induced symbolic system. In cognitive science, the idea has not been used as a means of proving theorems about complex systems, but rather as a way of simultaneously accounting for continuous and discrete aspects of cognition (which have historically been emphasized by opposed approaches to cognitive science). As Spivey ([2007]) has shown across a variety of domains—including categorization, language, visual perception, action, and reasoning—cognitive processes are often describable as both continuous *and* discrete. For example, subjects might assign a stimulus to one of a discrete set of categories, suggesting a mental process defined over a discrete set of representations. However, the *time it takes* the subject to make the assignment can vary depending on the stimulus and thereby reflect the operation of continuous dynamics (e.g., cats are more quickly categorized as mammals than whales are). Or, in cases where subjects make the assignments on a computer, the trajectory taken by the computer cursor sometimes reflects continuous dynamics (e.g. the cat icon is pulled straight to the mammal bin, while the whale icon is pulled briefly towards the fish bin before being pulled to the mammal bin).

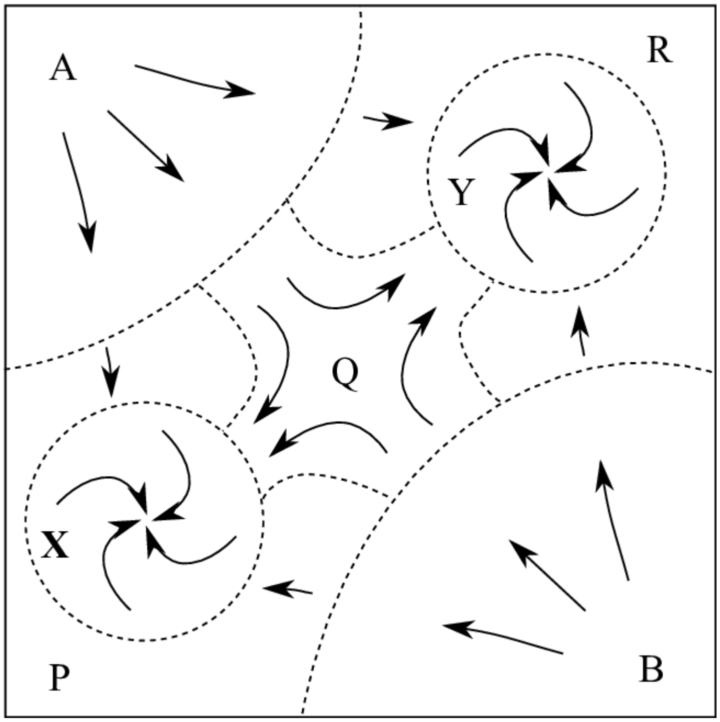

Figure 6. Idealized example of symbolic dynamics induced by a continuous base system. Adapted from Spivey ([2007], p. 112).

Dale and Spivey ([2005]) have suggested that symbolic dynamics could provide a useful way of conceptualizing such phenomena. For example, in figure 6, suppose B corresponds to a subject's state at the beginning of an experiment, and X and Y correspond to two possible categorizations. At the supervenient level, subjects will either move from B to Q to X, or from B to Q to Y (we are here leaving P and R out of consideration, and thinking of Q as an intermediate state of indecision). Though we have a discrete system—something like a probabilistic automaton defined over **S**—it reflects the dynamics of an underlying continuous system. For example, how long it takes the subject to move from B to X or Y depends on which orbit is being followed in the base dynamical system.

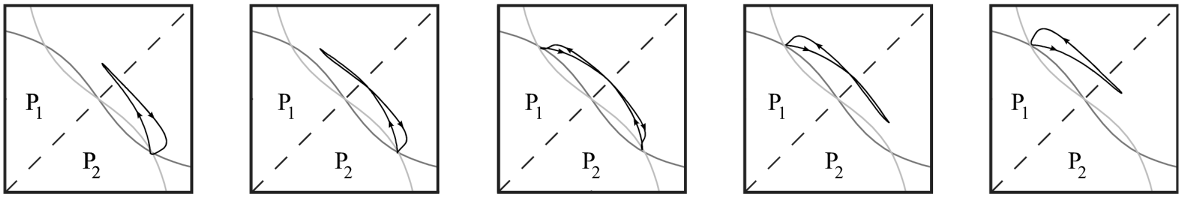

Figure 7: Behavior of a simple neural network relative to five environmental processes. The paths are shown in separate panels for convenience. The open phase portrait for this system contains these five paths, and many others, superimposed in the same state space.

The use of symbolic dynamics to conceptualize experimental results shows how supervenience functions and dynamical systems have already been usefully combined (at least tacitly) along the lines I suggested above. In Hotton and Yoshimi ([2010]), these ideas are made more explicit, with a special emphasis on *open* dynamical systems (see section 3). The general idea is that, by applying a supervenience function to an open phase portrait (which describes a system in an environment), it is possible to understand a variety of cognitive phenomena in a more detailed and mathematically tractable way than is otherwise possible. In figure 7, sample paths in an open phase portrait (a neural network model of an agent, embedded in a simple environment with two objects) are shown. The shape of each path reflects the way a particular environmental process interacts with the intrinsic attractor dynamics of the neural network.[xxx] The dotted line divides the state space into two regions, which correspond to two perceptual states for the agent: the region above the diagonal corresponds to a perception P_1 of the first object; the region below the diagonal corresponds to a perception P_2 of the second object. The base space in this model is $\mathbf{R}^2$, the supervenient state set is $\{P_1, P_2\}$, and the supervenience function f: $\mathbf{R}^2 \rightarrow \{P_1, P_2\}$ is defined as $f(X) = P_1$ if X is in the upper diagonal, P_2 otherwise. Using this function, each path in the open phase portrait of the neural network can be associated with a unique sequence of perceptual states. We have used this system to show how geometric tools can be used to study representational processes in embodied agents (so that the *shape* of a path reflects the way an agent perceives its environment over time), and to model various psychological phenomenon (for example, in the leftmost and rightmost figures the agent only fully perceives one object, though it is exposed to both; in those cases "masking" occurs, where one stimulus occludes another).

These examples give a sense of the utility of the overarching framework described in sections 1-3 for cognitive science, in particular for cases where a continuous system determines a discrete symbolic system. However, the framework is meant to be more general than that. In any case where a single system can be thought of in terms of levels related by supervenience, and where at least one of those levels has a dynamical system defined on it, the ideas above apply. For example, perhaps continuous neural dynamics determine continuous phenomenological dynamics. The nature of such a transformation—from the open dynamics of an embodied brain to the dynamics of consciousness—is something I hope to examine in future studies.

Acknowledgements

I am grateful to Rick Dale, Anthony Dardis, Scott Hotton, Stephanie Huette, Rolf Johansson, Victor Minces, Michael Spivey, Jessica Wilson, Cory Wright, and several anonymous referees, for helpful comments.

References

Abrahamsen, A. and Bechtel, W. [2006]: ‘Phenomena and mechanisms: Putting the symbolic, connectionist, and dynamical systems debate in broader perspective’, in R. Stainton (*ed*), *Contemporary debates in cognitive science*. Oxford: Basil Blackwell.

beim Graben, P., & Potthast, R. [2009]: ‘Inverse problems in dynamic cognitive modeling’, *Chaos: An Interdisciplinary Journal of Nonlinear Science*, **19**.

Bickle, J. [1998]: *Psychoneural Reduction: The New Wave*. Cambridge, MA: MIT Press.

Block, N. [1997]: ‘Anti-Reductionism Slaps Back’, *Philosophical Perspectives*, **11**, pp. 107-132.

Casey, M. [1996]: ‘The dynamics of discrete-time computation, with application to recurrent neural networks and finite state machine extraction’, *Neural Computation*, **8**, pp. 1135-1178.

Carnap, R. [2003]: *The Logical Structure of the World and Pseudoproblems in Philosophy*, Peru, Illinois: Open Court Classics.

Churchland, P. [1986]: ‘Some Reductive Strategies in Cognitive Neurobiology’, *Mind*, **95**, pp. 279-309.

Chalmers, D. [1994]: ‘On Implementing a Computation’, *Minds and Machines*, **4**, pp. 391-402.

Child, W. [1994]: *Causality, Interpretation, and the Mind*, Oxford: Oxford University Press.

Clark, A. [1998]: ‘Time and Mind’, *Journal of Philosophy*, **XCV**, pp. 354-376.

Clark, A and Chalmers, D. [1988]: ‘The Extended Mind’, *Analysis*, **53**, pp. 7-19.

Craver, C. [2007]: *Explaining the Brain*, Oxford: Oxford University Press.

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**, pp. 317-342.

Davidson, D. [1970]: ‘Mental Events’, in L. Foster and J.W. Swanson (*eds*), 1970, *Experience and Theory*, Amherst: University of Massachusetts Press, pp. 79-101.

Davidson, D. [1980a]: ‘Philosophy as Psychology’ in *Essays on Actions and Events*, Oxford: Oxford University Press.

Davidson, D. [1980b]: ‘The Material Mind’ in *Essays on Actions and Events*, Oxford: Oxford University Press.

Dardis, A. [2008]: *Mental Causation*. New York, NY: Columbia University Press.

Fodor, J. [1974]: ‘Special Sciences’, *Synthese*, **28**, pp. 97-115.

Gärdenfors, P. [2004]: *Conceptual Spaces: The Geometry of Thought*, Cambridge, MA: MIT Press.

Gurwitsch, A. [1964]: *The Field of Consciousness*. Duquesne, IL: Duquesne University Press,

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

Heil, J. [2003]: *From an Ontological Point of View*, Oxford: Oxford University Press.

Hellman, G. and Thomson, F. [1977]: ‘Physicalist Materialism’, *Noûs*, **11**, pp. 309-345.

Horgan, T. [1993a]: ‘From Supervenience to Superdupervenience’, *Mind*, **102**, pp. 555-586.

Horgan, T. [1993b]: ‘Nonreductive Materialism and the Explanatory Autonomy of Psychology’, in S. Wagner and R. Warner (*eds*), 1993, *Naturalism: A Critical Appraisal*, Notre Dame, IN: University of Notre Dame Press.

Horgan, T and Tienson, J. [2002]: ‘The Intentionality of Phenomenology and the Phenomenology of Intentionality’, in Chalmers, D. (*ed.*) *Philosophy of Mind: Classical and Contemporary Readings*, Oxford: Oxford University Press.

Hotton, S. and Yoshimi, J. [2010]: ‘The Dynamics of Embodied Cognition’, *International Journal of Bifurcation and Chaos*, **20**, pp. 1-30.

Izquierdo, E., Harvey, I., & Beer, R. D. [2008]: ‘Associative learning on a continuum in evolved dynamical neural networks’, *Adaptive Behavior*, **16**, pp. 361-384.

Kim, J. [1984]: ‘Concepts of Supervenience’, *Philosophy and Phenomenological Research*, **45**, pp. 153-176.

Kim, J. [1989]: ‘The Myth of Nonreductive Physicalism’, *Proceedings and Addresses of the American Philosophical Association*, pp. 31-47.

Kim, J. [2000]: *Mind in a Physical World*. Cambridge, MA: MIT Press.

Kim, J. [2003]: ‘Blocking Causal Drainage and Other Maintenance Chores with Mental Causation’, *Philosophy and Phenomenological Research*, **LXVII**, pp. 151-176.

Latham, N. [1999]: ‘Davidson and Kim on Psychophysical Laws’, *Synthese*, **118**, pp. 121-143.

Melynk, A. [2006]: ‘Realization and the Formulation of Physicalism’, *Philosophical Studies*, **131**, pp. 127-155.

McLaughlin, B. and Bennett, K. [2005]: ‘Supervenience’, in E. Zalta (*ed.*), *Stanford Encyclopedia of Philosophy* <plato.stanford.edu/>

Nagel, E. [1961]: *The Structure of Science*. New York: Harcourt, Brace and World.

Narens, L. [2002]: *Theories of Meaningfulness*. Mahwah, NJ: Lawrence Erlbaum Associates.

Noë, A. [2004]: *Action in Perception*. Cambridge, MA: MIT Press.

Pereboom, D [2002]: ‘Robust Nonreductive Physicalism’, *Journal of Philosophy,* **XCIX**, pp. 499- 531.

Oppenheim, P. and Putnam, H. [1958]: ‘Unity of science as a working hypothesis’, in H. Feigl, M. Scriven and G. Maxwell (*eds*), 1958, *Minnesota Studies in the Philosophy of Science*, **2**, pp. 3-36.

Shagrir, O. [1998]: ‘Multiple Realization, Computation, and the Taxonomy of Psychological States’, *Synthese*, **114**, pp. 445-461.

Shapiro, L. [2000]: ‘Multiple Realizations’, *The Journal of Philosophy*, **XCVII**, pp. 635-654.

Shoemaker, S. [2001]: 'Realization and Mental Causation', in Gillet, C. and Loewer, B. (*eds*), 2001, *Physicalism and its Discontents*. Cambridge: Cambridge University Press.

Shoemaker, S. [2007]: *Physical Realization*, Oxford: Oxford University Press.

Sober, E. [1999]: Physicalism from a probabilistic point of view. *Philosophical Studies*, **95**, pp. 135-74.

Spivey, M. [2007]: *The Continuity of Mind*. Oxford: Oxford University Press.

Van Gelder, T. [1995]: 'What might Cognition Be, if Not Computational?' *Journal of Philosophy*, **91**, pp. 345-381.

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

Wilson, J. [2005]: 'Supervenience-based Formulations of Physicalism', *Noûs*, **29**, pp. 426-459.

Yoshimi, J. [2007]: 'Supervenience, Determination, and Dependence', *Pacific Philosophical Quarterly*, **88**, pp. 114-133.

---

[i] However, in some cases—in particular, whenever I quantify over times—I tacitly separate the domain into a set of objects and a set of times; in those cases states are thought of as binary relations between objects and times.

[ii] We often say that an object is in "multiple states", but when we do those states are either of different types (a physical object at a time has mass, temperature, and velocity–but only one of each), or they are actually components of more complex states. Thus the three spatial coordinates of an object at a time correspond to a single, unique position in 3-space.

[iii] More specifically, a maximal property is produced by concatenating each atomic property (a property not formed from other properties) in a set of properties, or its negation, into one long conjunction, subject to the conditions that the resulting property is consistent and non-negative (i.e. it is not a conjunction of negated atomic properties).

[iv] In the phenomenological tradition a similar concept was studied by Aron Gurwitsch (Gurwitsch [1964]), who described "fields of consciousness" as "totalities of co-present data" (p. 3). In the analytic tradition Rudolph Carnap ([2003]) advocated beginning with what is "epistemically primary", namely, total mental states, what Carnap called "elementary experiences" (*erlebnisse*) or more famously, "the given", that is, "experiences themselves in their totality and undivided unity" (section 67).

[v] Relationships between abstract representations (e.g., numbers, or points in a state space) and the phenomena they symbolize are studied in measurement theory (see Narens [2002]).

[vi] I have argued elsewhere (Yoshimi [2007]) that extant conceptions of supervenience generally assume a "dependence" clause as well, whereby any object that instantiates a supervenient property instantiates some base property.

[vii] Latham ([1999]), calls these "supervenience conditionals." Latham is naming a concept found earlier in Kim ([1984]); also see the discussion of entailment in McLaughlin and Bennett ([2005]). What I call "determination" or "realization" relations has given rise to an expansive literature; see e.g., Bickle ([1998]), Shapiro ([2000]), Shoemaker ([2001]), Pereboom ([2002]), Melnyk ([2006]). I take the simple generalization I call "determination" to be consistent with most of these accounts. See note xii.

[viii] Note that F and G need not be of the same ontological category. This implies that supervenience, which corresponds to a family of realization relations, is neutral with respect to the mind-body problem (the point is generally acknowledged; see, e.g., Horgan [1993b]; Kim[2000]). In light of questions posed by qualia, I prefer to keep supervenience neutral with respect to the mind-body problem. However, if one prefers a strictly physicalist ontology, the realization relation can be suitably enhanced, without compromising the account given here.

[ix] More precisely: $\mathbf{B}' \subset \mathbf{B} = \{B \in \mathbf{B} \mid$ for all x, if x is in state B, there is some $A \in \mathbf{A}$ such that x is in state $A\}$. Note that $\mathbf{B}'$, so defined, need not be the domain of a supervenience function, since its members are not guaranteed to uniquely determine supervenient states. The proof of proposition 1 (which has been omitted due to space constraints) involves showing that *if* we assume **A** supervenes on **B**, then members of $\mathbf{B}'$ are guaranteed to uniquely determine supervenient states.

[x] Similar conceptions of supervenience have been articulated at various places in the literature. See, e.g., Kim ([1989]), Sober ([1999]), and Melnyk ([2006]), who refers to an earlier discussion of LePore and Loewer whereby explanation consists in systems of (in my terms) determination relations between physical and functional properties.

[xi] Briefly: strong local supervenience entails the same map described above, where determination of individual states is prefixed by a modal parameter (so that a base state X in $\mathbf{B}'$ determines the same supervenient state $f(X)$ in all worlds in the scope of the modal parameter), while weak local supervenience entails a family of supervenience functions, one for each possible world. Global supervenience induces a mapping from base states of worlds to supervenient states of worlds (strong global supervenience requires additional machinery to ensure that worlds instantiate relevant states in a manner which specifies which objects instantiate the local properties and relations that make up those states).

[xii] Those accounts typically add further components to the realization relation, e.g. requirements that ensure physicalistic acceptability (Melnyk [2006]), or the requirement that realizer properties differ in causally relevant ways (Shapiro [2000]).

[xiii] Also note that in the narrower scoped cases mathematical analysis of realization classes can yield philosophically interesting results. For example, one can literally measure the "dispersion" of realization classes in such cases (assuming a metric is associated with the base space), i.e. how much of the base space a realization class takes up and thereby assess multiple-realization arguments against reductionism using empirical evidence.

[xiv] The idea is familiar from multiple-realization based arguments against reductionism, according to which (1) reduction requires kind-to-kind bridge laws, but (2) realizers of

---

special science kinds are too heterogeneous to count as kinds. The classic discussion is Fodor ([1974]); on the relationship between similarity and kinds, see Block ([1997]).

[xv] A “supervenient dynamical system” or “supervenient system” is a dynamical system defined on a supervenient state set, and similarly for “base dynamical system” and “base system.”

[xvi] See Hasselblatt and Katok ([2002]), and Hotton and Yoshimi ([2010]).

[xvii] I will prefer the word “path”, which is non-standard but which is neutral between dynamical systems as classically defined and “open dynamical systems”, which account for a dynamical system’s interaction with an environment.

[xviii] If the supervenient state set consists of functional states, we can think of ϕ^r as an implementation of a functional architecture (via a straightforward extension of Chalmers [1994]). Also see Casey [1996], who describes how an important class of continuous dynamical systems (recurrent neural networks) can implement finite state machines.

[xix] The point can be seen informally by considering figure 4. Assume the base dynamical system is (unlike the example shown) continuous and consider arbitrary phase portraits on the base space. With few exceptions (e.g., the degenerate case where every path is a fixed point), these phase portraits violate the compatibility condition. The problem is even more acute in the case of figure 6. Insofar as these cases are relevantly similar to real-world cases, the compatibility condition will rarely be met in practice.

[xx] PROOF: Let $\phi^b : \mathbf{B'} \times \mathbf{T} \rightarrow \mathbf{B'}$ be a base dynamical system, and $f : \mathbf{B'} \rightarrow \mathbf{A}$ our onto-supervenience function. Let $\mathbf{Y} \in \mathrm{Pow}(\mathbf{A})$ be a set of supervenient states. Set $\phi^p : (\mathbf{Y} , t) := f(\phi^b(f^{-1}(\mathbf{Y}), \mathrm{t}))$. To show that ϕ^p is a dynamical system we must show: (1) $\phi^p(\mathbf{Y}, 0) = \mathbf{Y}$, (2) $\phi^p(\mathbf{Y} , t_1 + t_2) = \phi^p(\phi^p (\mathbf{Y}), t_1), t_2)$. (1) $\phi^p(\mathbf{Y},0) = f(\phi^b(f^{-1}(\mathbf{Y}), 0)) = f(f^{-1}(\mathbf{Y})) = \mathbf{Y}$. (2) $\phi^p(\mathbf{Y} , t_1 + t_2) = f(\phi^b(f^{-1}(\mathbf{Y}), t_1 + t_2)) = f(\phi^b(\phi^b(f^{-1}(\mathbf{Y}), t_1), t_2))$; $\phi^p(\phi^p (\mathbf{Y}), t_1), t_2) = f(\phi^b(f^{-1}(\phi^s (\mathbf{Y}, t_1), t_2))) = f(\phi^b(f^{-1}(f(\phi^b(f^{-1}(\mathbf{Y}), t_1))), t_2)) = f(\phi^b(\phi^b(f^{-1}(\mathbf{Y}), t_1), t_2))$. The proof assumes that functions can be applied to subsets of their domains to yield their images (for f and ϕ^b) or unions of their images (for f^{-1}).

[xxi] We are assuming for convenience that the state set and time set are finite. These assumptions could be relaxed, at the cost of a more cumbersome definition of degree of constraint.

[xxii] To see intuitively that collections of paths in a state space induce powerset dynamical systems, consider an initial bounding region **I** in a state space, and a future time *t*. Follow all points on all paths that pass through **I**, for a period of time *t*. The resulting set of points is the unique bounding region determined by the powerset system for initial condition **I** and future time *t*.

[xxiii] I take “autonomy” to designate a principle of disciplinary non-interaction, whereby practitioners of the autonomous discipline need not, in practice, consult lower-level sciences (even if the autonomous discipline studies properties which supervene on properties studied by those lower level sciences).

[xxiv] See Latham ([1999]) for discussion and further references. Also see Child ([1994]).

[xxv] Assuming a plausible environment; e.g. no mad scientist controlling my brain.

[xxvi] What remains to be done is to give a detailed analysis of how statistical laws can be derived from powerset dynamics, and to then use this analysis to derive one or more existing statistical laws in cognitive science from theoretical models of underlying neural activity.

[xxvii] A further worry is that the cases considered thus far emphasize an individual person's mind and brain, whereas explanation seems to apply to human psychology generally. However, the points made thus far apply to human psychology generally, if we consider base dynamics on a normalized brain space (e.g. the "Tailarach" space described in section 2). These dynamics would, by proposition 4, induce a powerset dynamical system on the space of human mental states, associated (I would argue) with general psychological laws.

[xxviii] Davidson seems to concede the point: "I take for granted that detailed knowledge of the neurophysiology of the brain will make a difference—in the long run, an enormous difference—to the study of subjects as perception, memory, dreams, and perhaps of inference" ([1980b], p. 247). But he thinks (wrongly, I have argued) that this sort of cross-disciplinary influence stops short of providing evidence of deterministic laws at the psychological level.

[xxix] See Dale and Spivey ([2005]); Spivey ([2007]); Harvey, and Beer ([2008]), and beim Graben, P., and Potthast ([2009]).

[xxx] The intrinsic attractor dynamics of the system are indicated by the pairs of curved lines in each frame, which are the two nullclines of the system, where the velocity in one direction is 0. The intersections of the nullclines are therefore fixed points, where the velocity in both directions is 0. Most states of the system are pulled towards these attracting fixed points.