TOPICS
TOPICS IN COGNITIVE SCIENCE

Topics in Cognitive Science 00 (2023) 1–16

ISSN: 1756-8765 online
DOI: 10.1111/tops.12699

This article is part of the topic "The Dynamical Hypothesis Three Decades Later: Advances, Critiques, and Prospects for a Dynamical Cognitive Science," Luis H. Favela and Vicente Raja (Topic Editors).

The Dynamicist Landscape

David L. Barack

Departments of Philosophy and Neuroscience, University of Pennsylvania

Received 15 March 2023; received in revised form 18 September 2023; accepted 19 September 2023

Abstract

The dynamical hypothesis states that cognitive systems are dynamical systems. While dynamical systems play an important role in many cognitive phenomena, the dynamical hypothesis as stated applies to every system and so fails both to specify what makes cognitive systems distinct and to distinguish between proposals regarding the nature of cognitive systems. To avoid this problem, I distinguish several different types of dynamical systems, outlining four dimensions along which dynamical systems can vary: total-state versus partial-state, internal versus external, macroscopic versus microscopic, and systemic versus componential, and illustrate these with examples. I conclude with two illustrations of partial-state, internal, microscopic, componential dynamicism.

Keywords: Cognition; Computation; Dynamical systems; Explanation

1. Introduction

Dynamical cognitive science is again on the rise. Dynamical systems in deep neural networks (Buckner, 2019; Krizhevsky, Sutskever, & Hinton, 2012; LeCun, Bengio, & Hinton, 2015) and in neuroscience (Barack & Krakauer, 2021; Barak, 2017; Favela, 2021; Sussillo, 2014; Sussillo & Barak, 2013; Vyas, Golub, Sussillo, & Shenoy, 2020) are increasingly used

Correspondence should be sent to David L. Barack, Departments of Philosophy and Neuroscience, University of Pennsylvania, 3700 Hamilton Walk, Richards Building A401, Philadelphia, PA 19104, USA. E-mail: dbarack@gmail.com

to understand the mind. But what is the relationship between dynamical systems and the mind?

What is a dynamical system? I assume there is a general ontological category of system that consists of every possible collection of objects, properties, and relations. Dynamical systems are systems that change over time or with respect to one another. The substrate of the dynamical system are the objects, properties, and relations that undergo change. The dynamics are those changes. Let "dynamical system" denote the objects, properties, or relations and the changes in them of some system.

These systems are described using dynamical systems theory, a branch of mathematics (Strogatz, 2001). Dynamical systems theory contains a set of concepts and tools for describing how systems change. A system's state space is the set of possible states for a system, where a state is the set of determinate objects, properties, and relations for the determinable types of the system. The dynamics of systems, the changes in systems over time, make up a series of states called a trajectory. Some trajectories tend to converge on points in the system's state space, called attractors, whereas other points repel trajectories, called repellors. These states are mathematically described by a set of values for all variables and parameters of the state equations describing the system's changes.

The seminal work of van Gelder provides an initial analysis of the relation between dynamical systems and minds (van Gelder, 1995; van Gelder, 1998; Van Gelder, 2006). van Gelder's dynamical hypothesis (DH) posits that "cognitive agents are dynamical systems" (Van Gelder, 2006, p. 73). I begin the discussion with a critical assessment of his (DH). Every physical system is a dynamical system of the sort defined by (DH), and so every cognitive agent is a dynamical system. More restricted classes of dynamical system can be specified by turning to four distinctions: total-state dynamical systems are contrasted with partial-state ones, microscopic with macroscopic ones, internal with external ones, and finally systemic with componential ones. Each type of dynamical system informs a type of dynamicism, a claim about what type of dynamical systems are cognitive as well as a specification of the modal scope of the claim (e.g., all, some, most cognitive systems are dynamical systems of such-and-so a type). My goal herein is not an exhaustive discussion but rather to highlight some distinctions for developing new dynamicisms. The intention is to outline different types of dynamical systems and so the distinctions are metaphysical. The conceptual geography complete, I present two examples that illustrate how partial-state, microscopic, internal, componential dynamicism, my preferred approach, can describe cognitive phenomena.

2. The dynamical hypothesis and the triviality challenge

The dynamical hypothesis (DH) maintains that cognitive agents are dynamical systems. As stated, however, the (DH) is trivial; every system including cognitive ones are systems that change over time or with respect to one another.[1] van Gelder proposes a more restricted (DH); nonetheless, his proposal still fails to distinguish between different views on the nature of cognitive systems.

van Gelder's (DH) is composed of two theses: the nature hypothesis and the knowledge hypothesis. I set aside the knowledge hypothesis for this discussion.[2] Turning to the nature hypothesis, van Gelder elaborates:

> "Cognitive agents instantiate numerous systems at any given time. …Second, cognitive agents 'are',… as many systems as are needed to produce all the different kinds of cognitive performances exhibited by the agent…. Another noteworthy fact about these models is that the variables they posit are not low level (e.g., neural firing rates) but, rather, macroscopic quantities at… the level of the cognitive performance itself…. [T]he nature hypothesis is concerned… with how agents are causally organized at the highest level relevant to an explanation of cognitive performances…. [T]he dynamical system responsible for a given kind of cognitive performance might include variables not literally contained within the agent itself" (van Gelder, 1998, p. 619).

In short, the nature hypothesis is that cognitive agents instantiate dynamical systems described by variables specified at the level of the cognitive phenomenon which may be both internal and external to the agent.

A few clarificatory points. First, the nature hypothesis states that the relevant relation between cognitive agents and their substrates is instantiation. Herein, I will assume that instantiation is (some version of) token identity. Second, van Gelder discusses cognitive "agents." To sidestep issues of agency, I will speak of systems in the sense outlined above.

Two distinctions in van Gelder's treatment will be important later. First, he distinguishes between low-level and high-level quantities, where low-level is below the level of the cognitive phenomenon and high-level is at that level. I assume these are spatiotemporal levels of description (cf. Churchland & Sejnowski, 1992), not levels of explanation (Barack & Krakauer, 2021), levels of analysis (Marr, 1982), levels of mechanism (Craver, 2015) or some other sense of level. I will dub this the macroscopic versus microscopic distinction. Second, he distinguishes between variables "literally contained within the agent itself" and those that are not in the agent. I will dub this the internal versus external distinction.

Considering the foregoing discussion, is there a statement of (DH) that focuses on the most contentious issues and sidesteps distractions? I propose the following statement:

> (DS): For every kind of cognitive phenomenon, there is some dynamical system instantiated by the cognitive system at the highest relevant level of causal organization such that that kind of cognitive phenomenon is a behavior of that dynamical system.

In (DS), I retain the idea in van Gelder's (DH) that cognition is a type of behavior and that cognitive systems instantiate dynamical systems, where the behavior is a behavior of that dynamical system.

Granted (DS) as the target for discussion, I will now argue that (DS) is trivial. For any cognitive system, there will be a dynamical system instantiated by the cognitive system at the highest relevant level of causal organization such that that kind of cognitive phenomenon is a behavior of that dynamical system. A dynamical system is some collection of objects, properties, and relations and changes in them. But every cognitive system is some collection of objects, properties, and relations and changes therein. Consequently, there will always be a dynamical system instantiated by the cognitive system at the highest relevant level of causal organization such that that kind of cognitive phenomenon is a behavior of that dynamical system.

How to defend against the triviality charge? While van Gelder acknowledges "some dynamical system or other" is always instantiated by a cognitive system, "[i]t is certainly not trivial that every cognitive performance is at the highest level a dynamical phenomenon. This is not true of ordinary digital computers" for example (van Gelder, 1998, p. 623). However, it is the case that every cognitive performance is at the highest level a dynamical phenomenon, as every such performance consists of objects, properties, and relations and how those change. Setting that aside, van Gelder continues by arguing that the triviality charge "…equivocates on the term 'dynamical system'. The [(DS)] takes cognitive agents to be dynamical systems in a much more specific sense, that is, quantitative systems" (van Gelder, 1998, p. 623). van Gelder continues: "[d]igital computers and dynamical systems are two classes of systems picked out by reference to different properties: roughly, effectiveness and interpretation as opposed to quantitativeness. …Turing machines bounce around their state spaces in ways [whose] order is based on formal properties, not quantitative properties" (van Gelder, 1998, p. 623–624). For van Gelder, it is the dependence on quantitative properties that is essential to dynamical systems: "a system is taken to be dynamical to the extent that it is quantitative" (van Gelder, 1998, p. 619) in one of the following senses: "(1) Quantitative in state. …a system is quantitative in state when there is a metric over the state set such that behavior is systematically related to distances as measured by that metric. …(2) Quantitative state/time interdependence. A system is quantitative in time when… there is a metric over the time set such that system behavior is systematically related to distances as measured by that metric. …(3) Rate dependence. …in some systems rates of change depend on current rates of change" (van Gelder, 1998, p. 618–619). In short, there are three senses of quantitative: in state, in time, or dependent on rate. A measure over the system, which must be "specifiable independent of system behaviour" (van Gelder, 1998, p. 655), is required to be quantitative.

van Gelder's proposal does rule out some dynamical systems from being cognitive ones, resolving the triviality charge. But the proposal fails to distinguish between proposals about the nature of cognitive systems, such as differentiating dynamical systems from digital computers. Digital computers are dynamical systems in the state quantitative sense (cf. Beer, 1998; Chater & Hahn, 1998).[3] All that is needed for such a measure is the following: for deterministic Turing machine T and constant rate (i.e., one state per time step), the distance between state S_1 and state S_2 is equal to the number of computational steps to get from S_1 to S_2. Behavior will then be systematically related to distance: for two behaviors starting from the same state, if one behavior occurs after the other, then more computational states occurred in the latter case than the former. Even if the behaviors are the same, the time of the behavior implies something about the cardinality of the set of states visited by the system, that is, the length of the system's trajectory through its state space. All that van Gelder requires is some systematic relation, and that is such a relation. Undoubtedly, this is a contrived measure, but regardless, something more is needed to demonstrate a sense in which cognitive systems are dynamical systems but digital computers (say) are not.

A second reply from van Gelder is that cognitive systems are dynamical systems "…at the highest relevant level of causal organization for a given kind of behavior. Digital computers do not satisfy this condition" (Van Gelder, 2006, p. 86). But digital computers do instantiate a dynamical system at the level of behavior because their behavior (e.g., changes in the text on a screen, the movements of some effector, and even input from, say, a keyboard) consists of changes of the objects, properties, and relations that make up the computer.[4]

van Gelder's appeal to quantitativeness restricts the class of systems that can be cognitive but fails to distinguish between different proposals about the nature of that class. Undoubtedly, there is some way to formalize the difference between dynamical systems like Watt governors and those like digital computers. For example, dynamical systems are typically described as continuous in time or state-space, whereas digital computers are not. Perhaps some way can be found to use this difference to infer a metaphysical difference between the two types of system. Instead of exploring novel senses of quantitativeness, the problems faced by van Gelder's proposal motivate finding other ways to distinguish between types of dynamical system. I turn to this project now.

3. Dimensions of dynamicism

My strategy is to offer novel resources for dynamicism by distinguishing types of dynamical system. I will specify four dimensions along which dynamical systems can vary: total-state versus partial-state, microscopic versus macroscopic, internal versus external, and systemic versus componential dimensions. My goal herein is to briefly describe these duals. The presentation of these distinctions is just a first step in developing novel dynamicist positions that are nontrivial and that may shine new light on the nature of cognitive systems. I hope that future work can further characterize the nature and interactions of the different dimensions.

First, total-state can be distinguished from partial-state dynamicism. Total-state dynamicism maintains that cognitive systems are the totality of their objects, properties, and relations (Clark, 1998; Port & van Gelder, 1995). Partial-state dynamicism maintains that cognitive systems may be some subset of the system's objects, properties, and relations. For example, consider deciding between two options, say choosing a blue square on the left and a red square on the right. A cognitive system that makes this decision will consist of some objects, properties, and relations, all of which are parts. These would include effectors like feet, hands, and other limbs; objects, properties, and relations that instantiate other cognitive functions; sensors like the eyes, ears, skin, and so forth; and more. On a total-state dynamicism, all these parts must be instantiated for a decision-making phenomenon to be instantiated. On a partial-state dynamicism, not all of these parts must be instantiated for there to be the cognitive phenomenon.

Do any philosophers endorse total-state dynamicism? Clark (1998) does, saying "[b]ecause it is assumed that there is… interanimation between multiple systemic factors…, the dynamicist chooses to focus on changes in total system state over time. The… models… thus reflect motion in a space of possible overall system states, with routes and distances defined relative to points, each of which assigns a value to all the systemic variables and parameters" (Clark, 1998, p. 364). Clark concludes his discussion by recommending "…the addition of an irreducibly dynamical dimension to the analysis" of mind (Clark, 1998, p. 375–376). Since this recommendation is couched in an understanding of dynamicism as total-states, he endorses a total-state dynamicism. What of partial-state dynamicism? In discussions of neural representation, some philosophers have implicitly sanctioned a partial-state description of dynamical systems underlying cognitive phenomena (see, e.g., Burnston, 2021a; Shagrir, 2012). I will briefly illustrate the partial-state approach with two case studies below.

Second, microscopic dynamicism can be distinguished from macroscopic. Macroscopic dynamicism describes dynamical systems with macroscopic properties, where macroscopic properties are those at the same spatiotemporal level of the cognitive phenomenon and microscopic ones are below that.[5] The distinction is carried over from van Gelder. A macroscopic dynamicism requires every property of the cognitive system to be at the level of the cognitive phenomenon; a microscopic one does not. Admitting microscopic properties provides a serious empirical boost to dynamicism since an appeal to these properties, notably including the mechanisms that give rise to the cognitive behavior, can distinguish between cognitive systems and noncognitive ones.

To illustrate the macroscopic versus microscopic distinction, consider synchronized finger movements, an example common in discussions of dynamicism (see, e.g., Haken, Kelso, & Bunz, 1985; Kelso, 1995). This example is typically described in terms of "macroscopic variables" (Haken et al., 1985) like the change in phase between finger movements. These "macroscopic variables" correspond to the macroscopic properties of the dynamical system. Because the phase of finger movements is described at the same level as the synchrony phenomenon, the phase is a macroscopic property. An alternative description in terms of "microscopic variables" of the nervous system and musculature can also be readily imagined; these would be microscopic properties.

Third, internal dynamicism can be distinguished from external. The external/internal division is relative to some boundary of the cognitive system; for my purposes, the skull or skin is a convenient dividing line. According to extended views of cognition, objects outside of the skin, such as phones, can still be part of the cognitive system (Clark & Chalmers, 1998). Events within the system and outside the system are linked in complex causal webs that can be taken to imply the importance of external properties for cognition (Clark, 1998). While an external dynamicism allows some properties external to the cognitive system to be essential to the cognitive phenomenon, an internal dynamicism does not.

Finally, fourth, systemic dynamicism can be distinguished from componential (Barack, 2019b). Systemic dynamical systems are those composed only of subsystems; typically, these take the form of brains/minds, bodies, and environments. Componential dynamical systems, in contrast, are those that permit both subsystems and subsubsystems, sub-sub-subsystems, and so on. The difference, then, is whether the components of subsystems are themselves parts of cognitive systems. A systemic dynamicism permits only subsystems; a componential dynamicism permits both subsystems and subsubsystems (and lower, i.e., sub…subsystems).

Consider again the above example of the cognitive system choosing between two options. On a systemic dynamicist view, the cognitive system is coupled to the decision environment, including the options and their associated outcomes. These can be thought of as a system composed of the option, the properties of the options like their colors, and the outcomes, which stand in a relation to their corresponding options. Often the decision environment is characterized by the difference in values between both options. How does the system change? There are a number of ways to conceptualize the dynamics underlying value-based decisions, but a recent approach models the agent as sampling the values of the options over time, integrating those samples over time, and then making a decision when the integrated evidence crosses a threshold (Krajbich & Rangel, 2011; Krajbich, Armel, & Rangel 2010). The state equations for the cognitive system include variables for the values of the options, implying that the two systems are coupled on the systemic approach. Alternatively, on a componential approach, there is a decision process that includes the values of the options but not the options themselves as a subsystem of the cognitive system.

While a full discussion of the different dimensions must await a different venue, I will briefly illustrate how the systemic versus componential distinction is orthogonal to the macroscopic versus microscopic one. The macroscopic versus microscopic distinction is anchored to the level of the cognitive phenomenon. Consider retrieving an item from memory, and let that cognitive phenomenon be described at the level of parts of the brain. Suppose the retrieval is the result of the interaction of brain networks (Ben-Yakov, Dudai, & Mayford, 2015; Nyberg, Cabeza, & Tulving, 1996), at the same spatiotemporal level as the retrieval phenomenon and so a macroscopic system. Now suppose the brain networks have parts such as brain areas or subnetworks that are at that same macroscopic scale and that are included in the dynamical system; then the dynamical system would be a componential one. Alternatively, the parts might be excluded, and so the dynamical system is a systemic one. Suppose instead the retrieval is the result of a synaptic mechanism (Aljadeff, Gillett, Obilinovic, & Brunel, 2021) at a lower spatiotemporal level than the retrieval phenomenon. This would be a microscopic system. The parts of the synapses that are at the same scale might be included in the dynamical system and so it would be a componential one; or, the parts might be excluded and retrieval explained due to the interaction of synapses, and so be a systemic one. The key is that the systemic versus componential distinction is defined in terms of parthood, whereas the macroscopic versus microscopic distinction is defined in terms of scale; as a result, the two distinctions are orthogonal.

Many philosophers who endorse dynamicism are systemic dynamicists (Beer, 1995; Beer, 2000; Chemero, 2011; Chemero & Silberstein, 2008; Port & van Gelder, 1995; Stepp, Chemero, & Turvey, 2011; van Gelder, 1995; Wheeler, 2005; Zednik, 2008; Zednik, 2011; Zednik, 2015). As van Gelder says, "…the true cognitive system is a single unified system embracing all three" of brain, body, and environment (van Gelder, 1995, p. 373). Because the "true" cognitive system is made up of all three subsystems, van Gelder (1995) endorses a systemic dynamicism. What of componential dynamicism? The distinction is drawn by Barack (2019b), who endorses a componential approach; such an approach is also discussed by a number of philosophers writing on the topic of the decomposability of dynamical systems (e.g., Bechtel, 1998; Burnston, 2021b; Kaplan, 2015; Zednik, 2011).[6]

Not every dynamical system instantiated by a cognitive system may fall under the same type of dynamicism. The nature of the relevant dynamical system could be relative to the cognitive phenomenon under discussion. Multiagent coordination to play a team sport may differ dramatically from paying attention to a stimulus. The former may be constituted only by multiple subsystems, a type of systemic dynamicism, while the latter need not. Insofar as this implies a sort of pluralist dynamicism across different cognitive phenomena, a pluralism is warranted.

The foregoing distinctions can have a range of impacts on our understanding of cognitive systems. First, they are meant to individuate different ontological types of dynamical system. Dynamicist claims have a metaphysical modal scope. For example, if cognitive systems are necessarily componential ones, then systemic dynamical systems, such as the system formed from the interaction of mind, body, and world as proposed by dynamicist proponents like van Gelder, are not the right kind of entity to be a cognitive system. The distinctions can also have methodological implications. If cognitive systems are usually componential ones, then some cognitive phenomena might be approached first with a componential dynamical system in mind; only after that assumption proves unfruitful would, say, a systemic dynamicism be adopted instead. The distinctions also suggest certain questions. For example, if cognitive systems are often but not always systemic ones, then componential dynamical systems might be possible but are relatively infrequent. But then why are they infrequent? Is this the result of some contingent fact about the way that cognition has evolved or manifests in the actual world? Or is it some deeper metaphysical truth about the nature of cognition that constrains their manifestation? The modal implications, the methodological implications, and the questions raised by dynamicist claims are important areas of debate that are all informed by drawing these distinctions. Most importantly, classifying dynamical systems using these categories will rule in some systems and rule out others, thereby providing new ways to distinguish between different proposals regarding the nature of cognition.

4. Partial, microscopic, internal, componential dynamicism

The fruitfulness of the foregoing distinctions will depend on the insight that they can provide for understanding cognition. As a brief illustration, I discuss two case studies from recent cognitive science: the neurocognitive mechanisms behind temporal interval estimation (Sohn, Narain, Meirhaeghe, & Jazayeri, 2019) and multitasking in recurrent neural networks (RNNs) (Driscoll, Shenoy, & Sussillo, 2022). I maintain that these two case studies from cognitive science are illustrations of partial, microscopic, internal, componential dynamicism. I have chosen these case studies because I contend that they are exemplary of much recent work in the neurocognitive sciences and that they illustrate a relatively overlooked type of dynamicism.

4.1. Temporal interval estimation

Consider first the investigation of mechanisms for estimating temporal intervals from Sohn et al. (2019). In the ready-set-go task, monkeys estimated an interval of time (Fig. 1A).

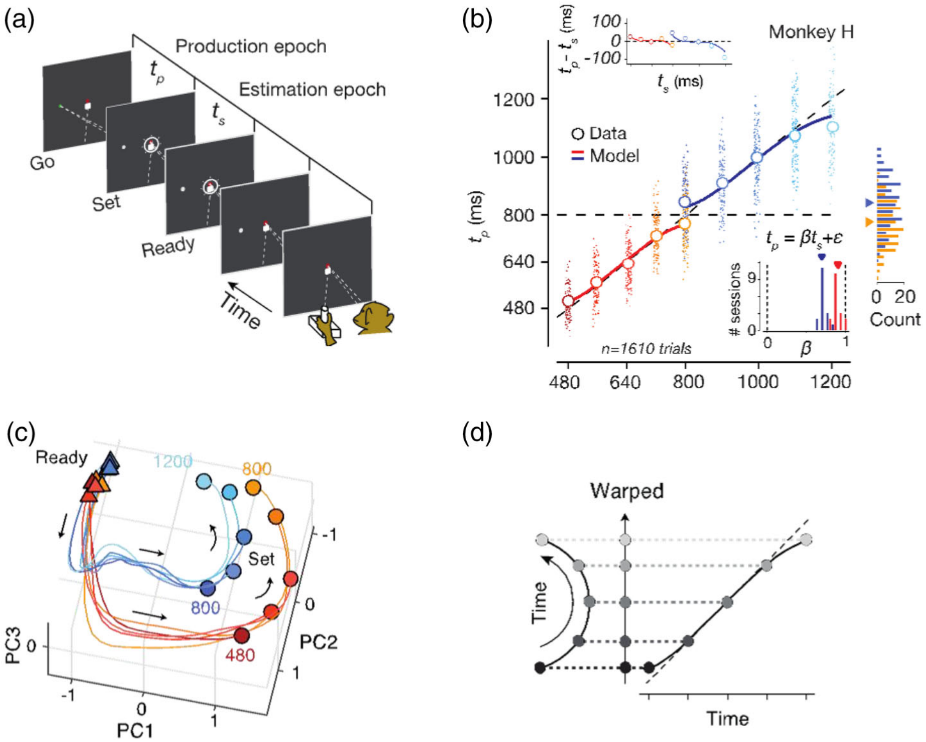

Fig. 1. The curved neural manifold and computational account for the temporal interval estimation task from Sohn et al. (2019). (A) The ready-set-go task for temporal interval estimation. (B) Behavior from the ready-set-go task. Monkeys showed a clear influence of the prior on their posterior estimates of intervals. (C) The projection of population activity during the estimation epoch into a 3-space derived from PCA during eye movements to the left for one monkey. Blue curve: Population activity from trials from the long interval distribution; orange curve: population activity from trials from the short distribution. Black arrows in C indicate the direction of time. (D) Illustration of how the projection of a curve on to a line warps the estimation of arc lengths along the curve. The effect of this warping is to pull the estimates closer to the mean of the distribution.

Source: https://www.sciencedirect.com/science/article/pii/S0896627319305628

Monkeys first saw a start cue ("ready") that indicated the start of the time to be estimated. Another cue ("set") signaled the end of this interval. The time from ready to set was the estimation epoch. Monkeys then estimated the elapsed time, after which they made a movement to a cue on a screen ("go"). Intervals were drawn from two different distributions, short and long, which were signaled by the color of the trial start cue. Notably, monkeys' estimates of intervals were skewed toward the means of the distributions (Fig. 1B).

To understand temporal interval estimation, neurons in the dorsomedial prefrontal cortex were recorded while monkeys completed the task. The neural activity was dimensionally reduced using principal components analysis (PCA) and projected into a low-dimensional space,[7] revealing a one-dimensional curved set of trajectories, or manifold, during the estimation epoch for each distribution of intervals (Fig. 1C). Different prior distributions map on to manifolds with different curvatures (the curvature of the longer distribution manifold was greater than the shorter distribution's manifold), which in turn map on to differences in the estimates of the intervals from each distribution (Fig. 1C). Further, differences in intervals map on to different points along the curved manifolds at the end of the interval, which in turn map on to differences in the estimates of the intervals within each distribution (Fig. 1C). Similarities in intervals from the same distribution map on to similar points along the curved manifolds, which in turn map on to similarities in the estimates of the intervals. Finally, for a given input interval and distribution, the skew toward the mean of the distributions could be explained by a linear read-out from the curvature of the manifold (Fig. 1D). In sum, the discovery of the manifold helps explain how monkeys estimate temporal intervals.

What about this case makes it relevant to dynamicism? The neural trajectories are defined in terms of changes in the location of the neural activity in the low-dimensional space. As they state, "firing rates for each [time interval] were estimated…. Neural trajectories… were analyzed within the subspace spanned by the top PCs that accounted for at least 75% of total variance. …In the estimation epoch, we examined the curvature in neural trajectories during the support of each prior by projecting" the neural state onto an axis in that low-dimensional space (Sohn et al., 2019, p. e3). These trajectories, then, are dynamical systems composed of firing rates computed from observed spiking in neurons, dimensionally reduced using PCA, and then projected into the lower-dimensional space.

What sort of dynamicism is supported by the findings from Sohn and colleagues? The work supports partial-state dynamicism, as it addresses only temporal interval estimation, ignoring other cognitive phenomena altogether. The objects and properties (such as the curved manifold) are below the organismic level of the cognitive phenomenon, temporal interval estimation, and as such supports microscopic dynamicism. The dynamics are changes in low-dimensional neural activity, which are internal to the system, and the analysis and modeling do not refer to other systems such as the body or environment, in support of a componential approach. In short, this case study of temporal interval estimation supports the conclusion that cognition is the result of partial-state, microscopic, internal, componential dynamical systems.

4.2. Multitasking in RNNs

The second case comes from the study of RNNs, neural networks part of whose output is input back into the network, a central tool to help understand how brains can instantiate cognitive systems (Sussillo, 2014; Barak, 2017). The basic state equation of an RNN is the standard firing rate equation formulated for a neural population (cf. Barak, 2017):

$$\tau \frac{d\boldsymbol{x}}{dt} = -\boldsymbol{x}(t) + \boldsymbol{W}\boldsymbol{x}(t) + \boldsymbol{B}\boldsymbol{u}(t),$$

where $\boldsymbol{\tau}$ is a vector of relaxation constants (one for each neuron in the network), $\boldsymbol{x}(t)$ is a vector of neuron activities at time t, $\boldsymbol{W}$ is a connectivity matrix specifying the weights from all neurons on to each other, $\boldsymbol{B}$ is a matrix of weights from input neurons on to the neurons in

$\boldsymbol{x}$, and $\boldsymbol{u}(t)$ is the activity vector for all the input neurons. This differential equation describes the change in activity in the RNN over time.

Driscoll and colleagues used RNNs to investigate how these networks learn to perform many different tasks (Driscoll et al., 2022). They trained the RNN on 15 different tasks, ranging from simple cued response tasks to delayed response tasks to delayed match to sample tasks. To understand how their networks were able to perform different tasks, they analyzed the RNN's activity by examining the network's location over time in the low-dimensional space that resulted from a PCA decomposition of the activity. In particular, they examined the trajectories around "fixed points" in this space, those points where the change in the RNN is zero. This revealed the presence of "dynamical motifs," "the high-dimensional nonlinear dynamics around a fixed point skeleton that implements computation for a specified input" (Driscoll et al., 2022, p. 3–4). In philosophy, Barack has called these "neurodynamical systems" (Barack, 2021). These dynamical motifs occurred for distinct phases of the task and are reused across tasks (Barack, 2019a).

To illustrate, consider the "MemoryAnti" and "MemoryPro" tasks. In the MemoryAnti task, the network was instructed to respond in a certain direction, waited through a memory timeout period, and then had to respond in the opposite direction. The MemoryPro task was similar, but the network had to respond in the instructed direction. The network's activity was projected into the first two dimensions of a PCA, and the trajectories during different epochs were examined. During the Memory period, when the network maintained the indicated direction in working memory, a ring attractor was observed for both MemoryAnti and MemoryPro tasks (Fig. 2A, left and right panels; MemoryAnti, purple; MemoryPro, yellow). Ring attractors are dynamical structure consisting of a ring-shaped series of attractors. This ring attractor was the same in both tasks (Fig. 2A, middle panel; smooth transition from purple to yellow), with distinct attractors around the ring for each stimulus direction the network kept in working memory. This explains how the RNN's activity maintains in working memory the different stimulus directions. A similar, shared ring attractor was observed during the Response period (Fig. 2B, middle panel), when the network generated a response, with distinct attractors for each response direction. Numerous such shared dynamical motifs were revealed in analyzing the RNNs, including across different RNNs, with different numbers of neurons, different activation functions, and more.

What sort of dynamicism is supported by these RNN findings? While they may seem to support a total-state dynamicism because Driscoll and colleagues constructed the network and specified its properties such as the number of units, the unit activation function, and the rule for learning network weights, further investigation revealed that "if the output of a particular cluster of units was set to zero, then all tasks involving a particular computation decreased their performance, while other tasks were unaffected" (Driscoll et al., 2022, p. 14). This suggests that only a subset of the network's properties is required for the cognitive phenomenon, supporting a partial-state view. The cognitive phenomenon resulted from the dynamics of regions in the low-dimensional space of the network: "Shared motifs are implemented by organizing the state in the appropriate region of state space to evolve on the relevant shared dynamical landscape" (Driscoll et al., 2022, p. 10), where "appropriate region" here means that "[t]asks with similar stimulus computations… organized their initial conditions for the

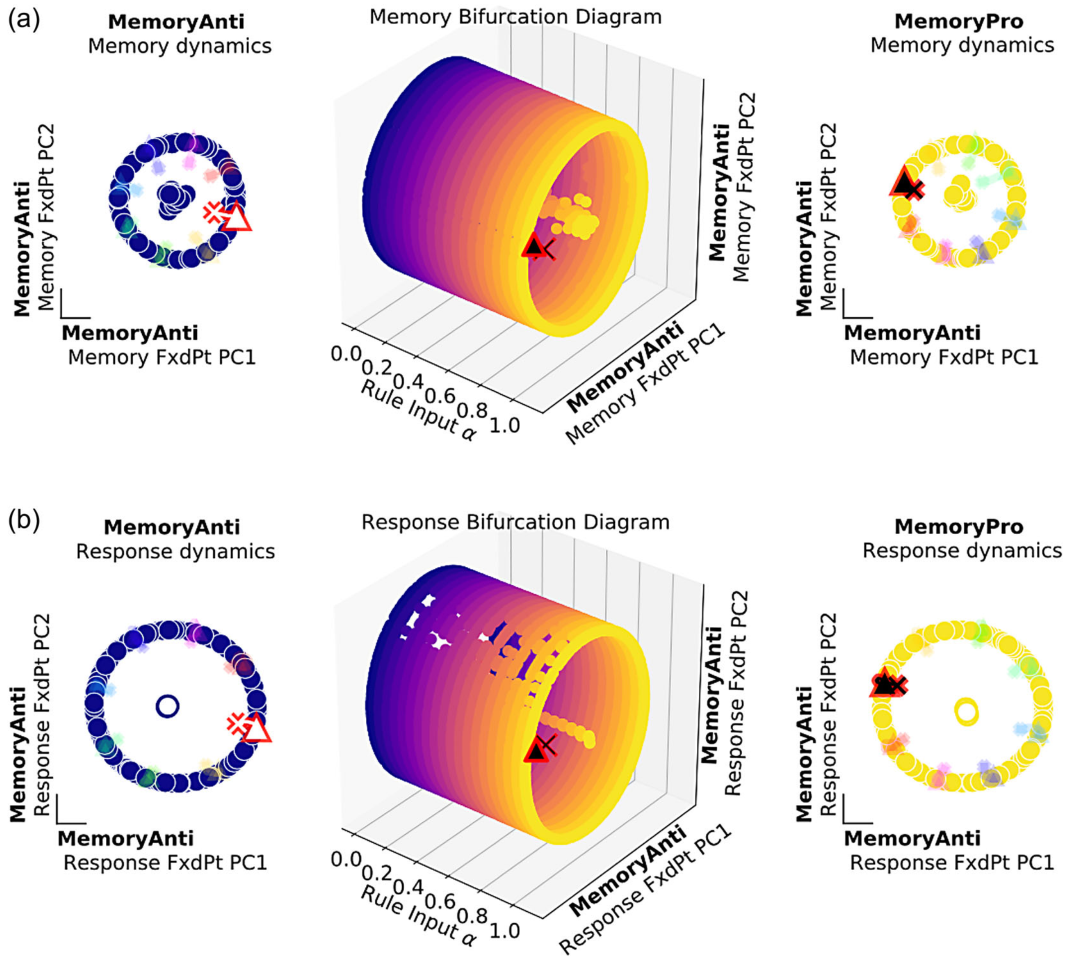

Fig. 2. Investigation of dynamical motifs in multitasking recurrent neural networks from Driscoll et al. (2022). (A) Microscopic dynamics for the memory phase of both the MemoryAnti task and the MemoryPro task. (B) Microscopic dynamics for the response phase of both tasks.

Source: https://www.biorxiv.org/content/10.1101/2022.08.15.503870v1.abstract

stimulus period to be nearby in state space and evolved in a similar way after stimulus onset" (Driscoll et al., 2022, p. 11). Hence, the activity in subspaces of the total activation space gave rise to the behavior of the whole network; insofar as the level of the cognitive phenomenon is determined by the network's behavior, the subspaces are below that level and so the microscopic properties of the network are included in the dynamical system. The dynamical system performing the cognitive task did not include properties outside of the network, supporting an internal dynamicism.[8] Finally, the case study implies the presence of components, whether those components are taken to be the dynamical motifs (as argued by Barack, 2019b) or the network neurons themselves. If the cognitive phenomenon is the behavior of the RNN coupled to a simulated environment, characterized in a way similar to the decision example above, then the RNN is a subsystem and the components would be subsubsystems, implying a componential dynamicism. The styles of analysis and the finding of shared dynamical motifs commend the view that the cognitive phenomenon in this case is the result of partial-state, microscopic, internal, componential dynamical systems.

5. Conclusion

Dynamicism in the neurocognitive sciences is alive and well—in fact, it is positively booming. While prescient, the formulation of the dynamical hypothesis by van Gelder applies to far too many systems and overlooks important distinctions between types of dynamical systems. Here, I described four duals: total-state versus partial-state; macroscopic versus microscopic; external versus internal; and systemic versus componential. These distinctions outline a landscape of dynamical systems, only some of which I explored herein, which can inform different dynamicisms. Different cognitive phenomena may be better described and explained by different types of dynamical system. Nonetheless, a type of dynamical system, one that is part of a larger system but that need not include it, that contains properties below the spatiotemporal level of description of the cognitive phenomenon, that consists of properties internal to the cognitive system, and that includes subsubsystems without requiring treatment of the body or environment, underlies some cognitive phenomena. I supported my claim with two case studies, one in cognitive neurobiology and the other from computational neuroscience. In short, at least some cognitive systems are partial-state, microscopic, internal, componential dynamical systems, illustrating the fruitfulness of the distinctions. Much more work on the nature of the distinctions is required. In particular, how the distinctions can be formally analyzed and whether they are all orthogonal remains to be determined. Nonetheless, I hope that laying out some of the space of types of dynamical system will lead to further refinements in our understanding of the nature of dynamical systems and in dynamical approaches to the mind.

Acknowledgments

Many thanks to Luiz Pessoa for the comments and to Luis Favela and Vicente Raja for encouragement.

Notes

1 This includes static systems if stasis is a type of zero change.
2 The knowledge hypothesis regards the nature of cognitive science. There is the nontrivial hypothesis that the nature of cognitive systems is, in part, determined by the nature of cognitive science. I set aside this possibility herein.

3 For an extended philosophical treatment of computational systems as dynamical systems, see Giunti (1997).
4 Philosophers have raised other problems for van Gelder's argument. For example, Piccinini notes that van Gelder's focus on descriptions of dynamical systems "…does not affect the nature of the system" (Piccinini, 2020, p. 250). Other critiques have been leveled (see, e.g., other commentaries attached to van Gelder, 1998 or in this volume).
5 To make good on this notion of levels, phenomena must be at some spatiotemporal scale. Notoriously, cognitive phenomena are often across-scales or scale-free (Barabasi and Albert, 1999; Bechtel, 2015; Marom, 2010). A scaled dynamicism must be supplemented to the extent that such scale-free phenomena are real and required for explaining cognitive phenomena.
6 Componential dynamicism is also compatible with representational and computational theories of mind. I endorse representational approaches, though I will not discuss them herein.
7 This is a dimensionality reduction technique in linear algebra; for details, see, for example, Strang (2016).
8 Unless the stimulus input and the rule signal, which dictated which task the network was to perform, are considered part of the system; but even then, this is not the typical sort of external property often alluded to in discussions of externalism (e.g., in Clark & Chalmers, 1998).

References

Aljadeff, J., Gillett, M., Obilinovic, U. P., & Brunel, N. (2021). From synapse to network: Models of information storage and retrieval in neural circuits. *Current Opinion in Neurobiology*, *70*, 24–33.

Barabási, A.-L., & Albert, R. (1999). Emergence of scaling in random networks. *Science*, *286*(5439), 509–512.

Barack, D. L. (2019a). Cognitive recycling. *British Journal for the Philosophy of Science*, *70*(1), 239–268.

Barack, D. L. (2019b). Mental machines. *Biology & Philosophy*, *34*(6), 1–33.

Barack, D. L. (2021). Mental kinematics: Dynamics and mechanics of neurocognitive systems. *Synthese*, *199*(1–2), 1091–1123.

Barack, D. L., & Krakauer, J. W. (2021). Two views on the cognitive brain. *Nature Reviews Neuroscience*, *22*(6), 359–371.

Barak, O. (2017). Recurrent neural networks as versatile tools of neuroscience research. *Current Opinion in Neurobiology*, *46*, 1–6.

Bechtel, W. (1998). Representations and cognitive explanations: Assessing the dynamicist's challenge in cognitive science. *Cognitive Science*, *22*(3), 295–318.

Bechtel, W. (2015). Can mechanistic explanation be reconciled with scale-free constitution and dynamics? *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, *53*, 84–93.

Beer, R. D. (1995). A dynamical systems perspective on agent–environment interaction. *Artificial Intelligence*, *72*(1), 173–215.

Beer, R. D. (1998). Framing the debate between computational and dynamical approaches to cognitive science. *Behavioral and Brain Sciences*, *21*(5), 630–630.

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

Ben-Yakov, A., Dudai, Y., & Mayford, M. R. (2015). Memory retrieval in mice and men. *Cold Spring Harbor Perspectives in Biology*, *7*(12), a021790.

Buckner, C. (2019). Deep learning: A philosophical introduction. *Philosophy Compass*, *14*(10), e12625.

Burnston, D. C. (2021a). Contents, vehicles, and complex data analysis in neuroscience. *Synthese*, *199*, 1617–1639.

Burnston, D. C. (2021b). Getting over atomism: Functional decomposition in complex neural systems. *British Journal for the Philosophy of Science*, *72*(3), 743–772.

Chater, N., & Hahn, U. (1998). What is the dynamical hypothesis? *Behavioral and Brain Sciences*, *21*(5), 633–634.

Chemero, A. (2011). *Radical embodied cognitive science*. MIT Press.

Chemero, A., & Silberstein, M. (2008). After the philosophy of mind: Replacing scholasticism with science*. *Philosophy of Science*, *75*(1), 1–27.

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

Clark, A. (1998). Time and mind. *Journal of Philosophy*, *95*(7), 354–376.

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

Craver, C. F. (2015). Mechanisms and emergence. In *Open MIND*, Open MIND. Frankfurt am Main: MIND Group.

Driscoll, L., Shenoy, K., & Sussillo, D. (2022). Flexible multitask computation in recurrent networks utilizes shared dynamical motifs. https://doi.org/10.1101/2022.08.15.503870.

Favela, L. H. (2021). The dynamical renaissance in neuroscience. *Synthese*, *199*(1–2), 2103–2127.

Giunti, M. (1997). *Computation, dynamics, and cognition*. Oxford University Press.

Haken, H., Kelso, J. S., & Bunz, H. (1985). A theoretical model of phase transitions in human hand movements. *Biological Cybernetics*, *51*(5), 347–356.

Kaplan, D. M. (2015). Moving parts: The natural alliance between dynamical and mechanistic modeling approaches. *Biology & Philosophy*, *30*(6), 757–786.

Kelso, J. A. S. (1995). *Dynamic patterns: The self-organization of brain and behavior*. Cambridge, MA, MIT Press.

Krajbich, I., Armel, C., & Rangel, A. (2010). Visual fixations and the computation and comparison of value in simple choice. *Nature Neuroscience*, *13*(10), 1292–1298.

Krajbich, I., & Rangel, A. (2011). Multialternative drift-diffusion model predicts the relationship between visual fixations and choice in value-based decisions. *Proceedings of the National Academy of Sciences*, *108*(33), 13852–13857.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). *Imagenet classification with deep convolutional neural networks*. In Advances in Neural Information Processing Systems.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444.

Marom, S. (2010). Neural timescales or lack thereof. *Progress in Neurobiology*, *90*(1), 16–28.

Marr, D. (1982). *Vision*. New York: Henry Holt and Co Inc.

Nyberg, L., Cabeza, R., & Tulving, E. (1996). PET studies of encoding and retrieval: The HERA model. *Psychonomic Bulletin & Review*, *3*, 135–148.

Piccinini, G. (2020). *Neurocognitive mechanisms: Explaining biological cognition*. Oxford University Press.

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

Shagrir, O. (2012). Computation, implementation, cognition. *Minds and Machines*, *22*(2), 137–148.

Sohn, H., Narain, D., Meirhaeghe, N., & Jazayeri, M. (2019). Bayesian computation through cortical latent dynamics. *Neuron*, *103*(5), 934–947. e935.

Stepp, N., Chemero, A., & Turvey, M. T. (2011). Philosophy for the rest of cognitive science. *Topics in Cognitive Science*, *3*(2), 425–437.

Strang, G. (2016). *Introduction to linear algebra*. Wellesley-Cambridge Press.

Strogatz, S. (2001). *Nonlinear dynamics and chaos: With applications to physics, biology, chemistry, and engineering (studies in nonlinearity)*. Westview Press.

Sussillo, D. (2014). Neural circuits as computational dynamical systems. *Current Opinion in Neurobiology*, *25*, 156–163.

Sussillo, D., & Barak, O. (2013). Opening the black box: Low-dimensional dynamics in high-dimensional recurrent neural networks. *Neural Computation*, *25*(3), 626–649.

Van Gelder, T. (1995). What might cognition be, if not computation? *Journal of Philosophy*, *92*(7), 345–381.
Van Gelder, T. (1998). The dynamical hypothesis in cognitive science. *Behavioral Brain Sciences*, *21*(5), 615–628.
Van Gelder, T. (2006). Revising the dynamical hypothesis. In P. Farias, & J. Queiroz (Eds.), *Advanced issues on cognitive science and semiotics* (pp. 73–91). Aachen: Shaker Verlag.
Vyas, S., Golub, M. D., Sussillo, D., & Shenoy, K. V. (2020). Computation through neural population dynamics. *Annual Review of Neuroscience*, *43*, 249–275.
Wheeler, M. (2005). *Reconstructing the cognitive world: The next step*. MIT Press.
Zednik, C. (2008). *Dynamical models and mechanistic explanations*. In Proceedings of the 30th Annual Conference of the Cognitive Science Society.
Zednik, C. (2011). The nature of dynamical explanation*. *Philosophy of Science*, *78*(2), 238–263.
Zednik, C. (2015). Heuristics, descriptions, and the scope of mechanistic explanation. In Pierre-Alain Braillard, & Christophe Malaterre (Eds.), *Explanation in biology* (pp. 295–318) Springer.