Synthese (2021) 199:1091–1123
https://doi.org/10.1007/s11229-020-02766-1

NEUROSCIENCE AND ITS PHILOSOPHY

Mental kinematics: dynamics and mechanics of neurocognitive systems

David L. Barack[1]

Received: 22 January 2019 / Accepted: 25 June 2020 / Published online: 1 July 2020

Abstract

Dynamical systems play a central role in explanations in cognitive neuroscience. The grounds for these explanations are hotly debated and generally fall under two approaches: non-mechanistic and mechanistic. In this paper, I first outline a neurodynamical explanatory schema that highlights the role of dynamical systems in cognitive phenomena. I next explore the mechanistic status of such neurodynamical explanations. I argue that these explanations satisfy only some of the constraints on mechanistic explanation and should be considered pseudomechanistic explanations. I defend this argument against three alternative interpretations of the neurodynamical explanatory schema. The independent interpretation holds that neurodynamical explanations and mechanisms are independent. The constitutive interpretation holds that neurodynamical explanations are constitutive but otherwise non-mechanistic. Both the independent and constitutive interpretations fail to account for all the features of neurodynamical explanations. The partial interpretation assumes that the targets of dynamical systems models are mechanisms and so holds that neurodynamical explanations are incomplete because they lack mechanistic details. I contend instead that the targets of those models are dynamical systems distinct from mechanisms and defend this claim against several objections. I conclude with a defense of the pseudomechanistic interpretation and a discussion of the source of their explanatory power in relation to a causal-mechanical description of the world.

Keywords Cognitive neuroscience · Explanation · Dynamical systems · Mechanisms

✉ David L. Barack
dbarack@gmail.com

[1] Presidential Scholar in Society and Neuroscience, Center for Science and Society, Departments of Philosophy, Neuroscience, and Economics, Columbia University, New York, USA

1 Introduction

Cognitive neuroscience describes both the dynamics and the mechanisms of the brain in explanations of cognition. While this dual nature reflects the explanatory and experimental endeavors of cognitive neuroscience, it also presents a conundrum for the philosophy of neuroscience. Are neural explanations of cognition mechanistic (Craver 2007b; Bechtel 2008; Kaplan 2011, 2015; Kaplan and Craver 2011; Piccinini and Craver 2011) or not (Port and van Gelder 1995; van Gelder 1995; Chemero and Silberstein 2008; Zednik 2011; Chirimuuta 2014, 2017; Huneman 2018)?

Recent philosophy of science has emphasized the mechanistic nature of many explanations across the sciences (Machamer et al. 2000; Bechtel 2002; Craver 2007b; Piccinini 2007; Glennan 2017). This mechanistic approach to explanation in science contends that phenomena are explained once they are situated in the causal-mechanical structure of the world (Salmon 1984). In contrast to the emphasis on mechanisms, dynamicism contends that some phenomena, including those related to cognition, are explained by a description of how properties of systems change over time or with respect to one another (Port and van Gelder 1995; van Gelder 1995; Chemero and Silberstein 2008; Walmsley 2008; Chemero 2011; Weiskopf 2011; Zednik 2011; Silberstein and Chemero 2012, 2013). The grounds for such explanations have been variously identified as deductive-nomological (Hempel and Oppenheim 1948), structural (Huneman 2018), optimal (Chirimuuta 2014), and more. I will generally refer to this diverse set of alternatives as the non-mechanistic approach.[1]

In this essay, I outline the central role that dynamics play in explanations of cognitive phenomena in cognitive neurobiology, the study of cognition by the investigation of activity in single and groups of neurons. After setting the stage by introducing dynamicism, I present a neurodynamical explanatory schema that captures the explanatory role of neurodynamical systems and illustrate this schema with a case study. I consider four different interpretations of the schema. The independent interpretation maintains that the schema outlines non-constitutive explanations and is independent of mechanisms altogether. The constitutive interpretation maintains that completing the schema results in constitutive but otherwise non-mechanistic explanations. The partial and pseudomechanistic interpretations both agree that neurodynamical explanations are constitutive explanations that situate cognitive phenomena in the world's causal-mechanical structure. The partial interpretation further maintains that dynamical systems theoretic models in these explanations target mechanisms but are incomplete descriptions of them and, so, the explanations that result from the completion of the schema are also incomplete. In contrast, the pseudomechanistic interpretation maintains that dynamical systems theoretic models in those explanations target non-mechanistic neurodynamical systems.

[1] Of course, one could adopt a pluralistic attitude toward cognitive neuroscientific explanations, where some explanations are mechanistic and some not.

I argue for three main claims in this essay. First, the neurodynamical explanatory schema is a constitutive, productive explanatory schema, ruling out both independent and constitutive interpretations. Second, some dynamical systems theoretic models in cognitive neurobiology target neurodynamical systems and are not incomplete models of mechanisms. Third, neurodynamical systems are not mechanisms except perhaps in a novel sense of the term and so neurodynamical explanations are pseudomechanistic explanations. Neurodynamical systems are often described in a fashion that is independent to some extent of the underlying neurophysiological mechanisms that instantiate them. These neurodynamical systems comprise the kinematics of mind.

2 Cognitive neurobiology and neurodynamics

In this section I illustrate the claim that dynamical systems constitute cognitive systems and present a neurodynamical explanatory schema built on this dynamical foundation. I first briefly discuss dynamical systems and dynamical systems theory, a branch of mathematics used to describe how systems change. I then outline a neurodynamical explanatory schema that will serve as an anchor point for the rest of the discussion. I illustrate this schema with an example drawn from the neuroscience of perceptual decision making.

2.1 Dynamical systems theory

I assume that there is a general ontological category of system that consists of (more or less) every possible collection of objects, properties, and relations.[2] Dynamical systems are systems that change over time or with respect to one another. There are two elements in a dynamical system, the substrate and the dynamical properties. The substrate of the dynamical system are the objects, properties, and relations that undergo change. The dynamical properties are the changes in the substrate. Let 'dynamical system' denote the objects, properties, or relations of some system as well as the changes in the objects, properties, or relations in that system.

These systems are described using dynamical systems theory, a branch of mathematics (for extended introduction, see Strogatz 2001). Dynamical systems theory contains a set of concepts and tools for describing how systems change. A system's state space is the basic concept in dynamical systems theory, the set of all possible states that a system can occupy. A state for such a system will be defined as the set of determinate objects, properties, and relations for the determinable types of the system. These states are mathematically described by a set of values for all the variables and parameters of the equations that describe the system's change, the system's state equations. These state equations are either difference equations,

[2] Object here is used in a very general sense. Also, a system need not include elements from all three categories; a system could be a collection of properties only, for example.

describing discrete change in these variables, or differential equations, describing continuous change.

As systems evolve over time (or with respect to some other variable), they inhabit a series of states drawn from their state space. This series of states is called a trajectory. A system's state evolution is a description of the system in terms of the trajectories taken and is mathematically described by the state equations. The system's state space can possess structure, such as when trajectories tend to converge on or near a single state, called an attractor. Other structural features can be present, including limit cycles, stable repeating sequences of states; bifurcations, where small changes in state yield large changes in trajectories; repellers, states that trajectories move away from; and more. Often the structure in its state space determines the type identity of a dynamical system.

2.2 Neurodynamical systems and dynamical systems models

Neurophysiological systems are organized collections of neural objects, properties, and relations. These systems are constantly changing across spatial and temporal scales. Neurodynamical properties are the changes in the objects, properties, or relations of neurophysiological systems. For the following, neurodynamical systems are organized collections of these neurodynamical properties and their substrates.[3] Since these dynamical properties are properties of neurophysiological systems and their substrates are identical to objects, properties and relations of such systems, neurodynamical systems are token identical to subsets of neurophysiological systems.[4]

The approach herein is distinct from other approaches to dynamical systems in discussions in the philosophy of neuroscience. A key distinction is between dynamical systems theoretic (DST) models and dynamical systems (cf. Giunti 1997; van Gelder 1998; Beer 2000). Models are representations that have targets (Weisberg 2013; Glennan 2017), and different models can have the same or different targets. This distinction between DST models and dynamical systems is often left out of discussions of dynamicism. For example, Shapiro states straight away that "dynamical systems are mathematical models of real world systems that exhibit change" (Shapiro 2013, p. 354). Similarly, Lyre states that "[d]ynamical systems consist of states given as sets of variables that can be represented by points in a state space where the evolution function is usually the solution of a (system of) differential equation(s)" (Lyre 2017, p. 4).[5] Both of these statements violate the distinction between DST models, which are mathematical models, and dynamical systems themselves, which are sets of dynamical properties and their substrates, not mathematical models. The

[3] I have specifically used the phrase 'organized collections' because cognitive neurobiologists refer to systems that feature some temporal or topological structure. However, in principle, the collections need not feature this structure.

[4] The dynamical properties are themselves subsets of the set of all properties of such systems. Neurodynamical systems are typically proper subsets of dynamical properties of neurophysiological systems because usually not every dynamical property of such a system gets included in the neurodynamical system.

[5] I thank an anonymous reviewer for soliciting this comparison to Lyre.

models target those systems. The nature of the substrate of a dynamical system dictates how much spatiotemporal or physical detail to include in its description (cf. Weiskopf 2017) on the distinction between cognitive models and mechanistic ones). Some philosophers reject the possible inclusion of such detail. For example, Lyre restricts dynamical systems to describing structure, "…a set of relations imposed on a set of relata such that the relata are only defined via the relations in which they stand" (Lyre 2017, p. 7). On Lyre's account, dynamical systems are described by models whose equations refer only to relations. My approach, in contrast, includes all changes, whether in objects, properties, or relations.[6] Take, for example, a neuron that discharges an electrical pulse or action potential (i.e., a firing neuron). On the view herein, the objects, properties, and relations underlying this firing, such as the ions, proteins, and so forth, are distinguished from the changes in those constituents that constitute the firing. A neuron fires when a sudden and large change in the cell membrane voltage propagates down the cell. These changes are the neurodynamical properties of the neuron, and the collection of those changes and their substrate—the ions, proteins, and so forth—constitute a neurodynamical system. But the firing rate of a neuron can also be modulated. In that case, the neurodynamical properties are changes in the firing rate of the neuron, and the neurodynamical system consists of those changes together with their substrate, the firing rate itself. Neurodynamical systems can refer to changes in neural objects, properties, or relations or changes in changes in those objects properties, or relations, or changes in the changes in the changes, and so forth.

2.3 Explanation in cognitive neurobiology

The neurodynamical explanatory schema contains six steps. The explanandum is some cognitive capacity. First, this cognitive capacity is decomposed into a set of subcapacities, functions performed by the subsystems of a system for the system, that in concert yield the cognitive capacity.[7] Second, one of these subcapacities is selected for analysis. Third, this subcapacity is described mathematically. Fourth, the dynamics of the neural subsystem, the one hypothesized to possess the subcapacity, are described. Fifth, those dynamics are also described mathematically. Sixth and finally, the mathematical description of the subcapacity is mapped on to the mathematical description of the dynamics. The conclusion is that the dynamical system has the subcapacity, that is, that the dynamical system performs a function for the cognitive system. The neurodynamical system helps explain the cognitive capacity in virtue of possessing the subcapacity.

[6] On some views, all properties, or all scientific or physical properties, are relational. On such views, Lyre's approach and my own will collapse with regard to this second point.

[7] Of course, to fully explain a cognitive phenomenon, each subcapacity must be accounted for. I skip over this complication in the following discussion.

I will briefly review a case study of perceptual decisions under noisy sensory conditions from cognitive neurobiology to illustrate the schema.[8] Neuroscientists study these decisions using the random dot motion task (RDMT), in which subjects are presented with a visual display of moving dots and decide on the direction of motion. Some fraction of the dots moves in the same direction, and different fractions of dots move coherently on different trials. The coherent motion signal is embedded in random dot noise. Subjects indicate their decision by looking at a target. A number of functions are required for the RDMT: stimulus detection, motion encoding, integration of motion evidence, option selection, motor selection, motor initiation, and so forth.[9] For evidence integration, a sequential probability ratio test (SPRT Wald and Wolfowitz 1948) describes an evidence sampling process for determining the direction of motion. In the SPRT, the prior odds of the dots moving left or right are set first. Next, motion evidence from the field of moving dots is gathered. After evidence is gathered, the odds that the dots are moving left or right are updated on the basis of that evidence. This evidence gathering and updating process continues until a decision threshold is crossed.

In primates, motion properties of visual stimuli are encoded in area V5/MT, an area in the occipital cortex of the brain (Zeki 1974, 1991; Britten et al. 1992), and communicated to the lateral intraparietal area (LIP; Blatt et al. 1990), an eye movement control region in the parietal cortex (Britten et al. 1992, 1993; Platt and Glimcher 1999). An organized set of dynamical properties, the integrate-to-bound system, for integrating this motion evidence is found in area LIP (Roitman and Shadlen 2002; Gold and Shadlen 2007). The integrate-to-bound dynamical system starts at an initial state, transitions through a series of adjacent states, and then terminates at the same point across different initial states and trajectories. The system continuously changes state as a smooth, non-saltatory function of changes in environmental or internal properties until a boundary is reached. After arriving at the terminal state, the system resets to the initial state.

The state equations for the integrate-to-bound system uncovered in LIP neuronal activity are described using a variety of mathematical formulae (Usher and McClelland 2001; Wang 2002; Mazurek et al. 2003; Ditterich 2006; Wong and Huk 2008), many of which are ordinary differential equations. The simplest models of neural integrators describe them as instances of leaky integrators with feedback (cf. Goldman et al. 2010), p. 167):

$$\tau_{neuron}\frac{dr}{dt} = -r + wr + I(t)$$

[8] See Gold and Shadlen (2007) for extensive discussion of this research. Note that many aspects of this case are still actively researched and hotly debated (Latimer et al. 2015; Shadlen et al. 2016). For my present purposes, the still unsettled details do not matter, as I am merely illustrating how such explanations are constructed.

[9] Different analyses of this task will yield different sets of functions; the specific set selected does not matter for the moment.

for neuron time constant τ_{neuron}, firing rate r, synaptic strength w, and input current I(t).[10] The first term on the right hand side represents the decay in firing rate, the second term represents the weighted feedback into the system, and the third represents the input to the system. This simple differential equation describes how the firing rate changes as a function of input and feedback to the neuron. The key feature is that the input I is integrated over time. Under constant input, the integrator will increase in firing rate in proportion to the input, feedback, and leak. This equation is complicated in various ways in order to take into account types of feedback, baseline firing rate effects, and maximum firing rate effects. The integrate-to-bound system contains two key additional features. First, upon reaching a particular threshold, the firing rate resets to some baseline value. Second, the same threshold exists for different strength inputs. There are different ways of mathematically incorporating these features (see, e.g., Wong and Wang 2006).

In dynamical systems theory, for a single input this integrate-to-bound system is called a one-dimensional attractor with a single fixed point: the system is drawn toward one point in its state space (Strogatz 2001; Goldman et al. 2010). The integrate-to-bound system will transition through the space of firing rates to a particular value and then reset. This particular value, the threshold, is an attractor in the system's state space. The threshold is an unstable attractor; if the system moves beyond the threshold, it resets to the baseline state. For different inputs, the integrate-to-bound system has the same fixed point attractor. The integrative activity of the integrate-to-bound system across different inputs can be depicted using one dimension for the space of possible activity patterns (such as the set of possible firing rates) and another for the different possible inputs (Fig. 1b). For each input, the system is drawn to the same firing rate, tracing a line in the state space composed of the two dimensions of the changes in firing rates and the possible inputs. This is called a line attractor in dynamical systems theory. The integrate-to-bound system behaves differently for the same firing rate and input depending on whether the system is pre- or post-threshold. One way to depict the full integrate-to-bound system uses a disjoint state space for pre- and post-threshold activity. For pre-threshold activity, the system integrates (Fig. 1b, below threshold). For post-threshold activity, the system returns to a baseline firing rate regardless of the input (Fig. 1b, above threshold).

This case instantiates the neurodynamical explanatory schema. The explanandum is the capacity to make perceptual decisions about the direction of motion in noisy conditions. First, the cognitive capacity for motion discrimination is decomposed into a set of subcapacities like motion processing, evidence integration, and so on. Second, the relevant subcapacity, namely evidence integration, is specified. Third, this evidence integration function is described mathematically with the SPRT. Fourth, the integrate-to-bound dynamics of the physical subsystem for motion discrimination are described as an integrating trajectory through the subsystem's state space towards an attractor point. Fifth, those dynamics are described mathematically with the specification of a state equation that is a function of time and

[10] This equation is often used to describe the firing rate of a pool of neurons. Here I use it to describe a single neuron.

motion evidence. Sixth and finally, the mathematical description of the subcapacity is mapped on to the mathematical description of the dynamics. In particular, some range of the mathematical description of the SPRT maps on to some range of the mathematical description of the integrative activity observed in area LIP.[11] The conclusion is that the integrate-to-bound system integrates motion evidence for making perceptual decisions.[12]

In what follows, I will assume that the neurodynamical explanatory schema has explanatory power—that is, that completing the schema results in an explanation.[13] A number of philosophers have argued that mechanisms are central to explaining cognitive phenomena. This leads to the main question for the remainder of this essay: how do neurodynamical systems and neurodynamical explanations relate to mechanisms and mechanistic explanations?

3 Mechanistic explanation

In the foregoing, I described neurodynamical systems as neurophysiological systems and their changes and illustrated how neurodynamical systems play a fundamental role in explaining cognitive phenomena. Mechanists argue that explanations in cognitive neuroscience proceed by describing the mechanisms underlying cognitive phenomena. In order to evaluate the mechanistic claim in light of the neurodynamical explanatory schema, I must discuss mechanistic explanation.

I analyze mechanistic explanations into three conditions. First, mechanistic explanations are constitutive explanations, explanations that appeal to some system that constitutes the explanandum (Salmon 1984; Craver 2007a, b). Second, the system is a mechanism in a technical sense: an organized collection of entities and activities that help to produce some outcome. Third, the system produces the explanandum phenomenon, ensuring that the occurrence of the explanandum is the result of the system. These three conditions are jointly sufficient for a mechanistic explanation.

3.1 Constitution

Mechanistic explanation is constitutive explanation that situates the explanandum in the causal-mechanical structure of the world. As Craver puts it, "causal-mechanical explanations explain by showing how something is situated within the causal nexus" (Craver 2007a, p. 4). A mechanistic explanation describes how a system's parts and organization makes up the explanandum. The first condition on

[11] While I do not have space to delve deeper into this mapping, note that a 1:1 mapping within a range is insufficient.

[12] Recent research has called into question whether area LIP in fact causally influences evidence integration during motion discrimination (Katz et al. 2016). Nonetheless, this example serves to illustrate the explanatory structure.

[13] The defense and scope of the schema are presented elsewhere (Barack 2019), where I contend that the schema has broad application in cognitive neurobiology.

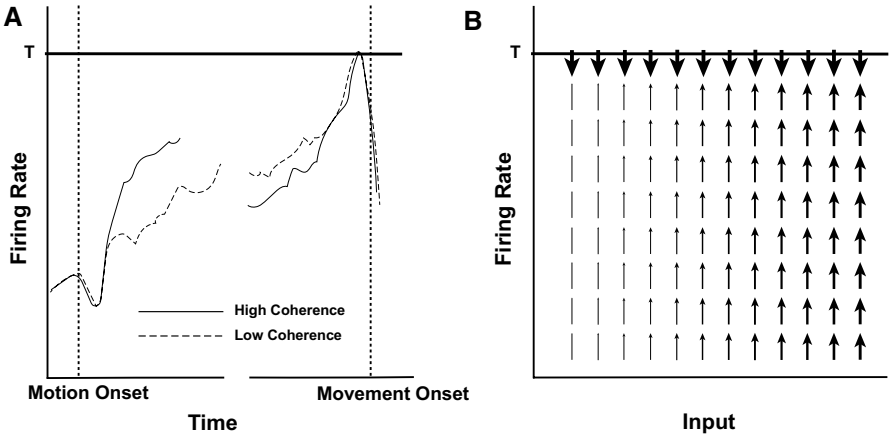

Fig. 1 **a** Depiction of average recorded firing rates from a cell in the parietal cortex of a monkey while they make a perceptual decision in different evidential conditions. Time is on the x-axis and firing rate is on the y-axis. Solid line is high motion coherence (strong evidence) and dashed line is low motion coherence (weak evidence). First dotted vertical line corresponds to motion onset and second to movement onset. A larger proportion of dots moving in the same direction results in a steeper increase in activity of neurons. After Roitman and Shadlen 2002, p. 9482. **b** A schematic depicting the state space for this neuron. Arrows originate at points in the state space, and only a subset are shown for clarity. Arrow orientation indicates the direction in which the system moves through its state space and arrow thickness the size of the change in firing rate under a given input (strength of evidence). As input is received by the neuron, firing rate increases, and the system's trajectory passes through a series of firing rates. Upon reaching the threshold T, the system resets to the baseline firing rate, as indicated by the row of downward-pointing arrows that originate just past the threshold. This depiction combines the pre-threshold state space (below T) and the post-threshold state space (above T) in the same diagram

mechanistic explanation is an organized system with parts that constitutes and situates the explanandum in the causal-mechanical structure of the world.

3.2 Mechanisms

Mechanistic explanations explain some phenomenon by referring to a mechanism. The first condition suggests a general definition of a mechanism as a productive, organized set of parts that play roles (cf. Levy and Bechtel 2016). On this general definition, any collection of objects that have properties and stand in relations that appears in an explanation—that is, any system—may be a mechanism. This general definition of a mechanism, however, may reasonably be charged with triviality. If any set of organized parts that play roles that appears in an explanation is a mechanism, then the scope of what counts as a mechanism is overly large (Campbell 2008, p. 430; Dupré 2013, p. 28; Franklin-Hall 2016; Chirimuuta 2017, p. 22ff; Paz 2017, p. 220ff; Weiskopf 2017, pp. 56–57). Without further restrictions, any system might be a mechanism.

The recent turn to mechanism in the philosophy of science presents a different definition of a mechanism. Two brief examples illustrate this definition. Consider

the heart, an organ that circulates blood throughout the body.[14] The heart is composed of a set of parts, such as an aorta, valves, ventricles, cellular tissue and other structures, with associated things that the parts do, like contracting rhythmically or preventing backflow, organized such that the heart pumps blood received from the body to the lungs and from the lungs through the aorta to the body. As another example, consider the biophysical mechanism of a chemical synapse, such as a voltage-sensitive Na^+ channel.[15] The mechanism consists of a set of parts, including the cell membrane, vesicles, microtubules, ions, etc. and a set of things that the parts do, including biosynthesis, coupling, diffusion, transport, depolarization, etc. These parts and what they do are organized such that the ion channel opens in the presence of spreading depolarization along the cell, allowing influx of ions and the propagation of the depolarization. In both cases, the explanation consists in an organized set of parts and their roles such that they produce the explanandum phenomenon, such as propagating an electrical signal or pumping blood.

These two examples illustrate how the recent approach in the philosophy of science emphasizes physical, machine-like properties (Machamer et al. 2000, p. 3ff; Bechtel and Abrahamsen 2005, p. 423ff; Wright and Bechtel 2007, p. 45ff; Piccinini 2010, p. 285ff; Illari and Williamson 2012, p. 119ff). Mechanisms can then be defined as:

> m-mechanism $\stackrel{\text{def}}{=}$ an organized set of parts characterized by their physical properties (such as spatiotemporal location, shape, electrochemistry, etc.) and that fill roles characterized by their physical properties (such as rate, duration, material changes, etc.) that causally produce and are responsible for some phenomenon.

The parts of a m-mechanism are individuated on the basis of their physical properties (spatiotemporal, electrical, etc.) and their causal role in producing the phenomena by being oriented, structured, etc. in the right fashion. In keeping with current lingo, the parts of systems that satisfy the definition of a m-mechanism are entities. The roles of the m-mechanism are what the entities or their interactions do in the m-mechanism. The things that entities do to fill roles are activities. These activities are defined by their temporal order, rate, and duration, and individuated by the entities that engage in them or their spatiotemporal properties.[16] The entities and activities of the m-mechanism are organized so as to allow for the production of the explanandum. As is widely recognized, m-mechanisms are also functionally individuated, in the sense of being responsible for a phenomenon (the so-called

[14] Bechtel and Abrahamsen (2005, p. 424ff) discuss the heart as mechanism. In general, Bechtel and colleagues are more liberal in their approach to mechanisms than Craver, Kaplan, Piccinini and colleagues, and so are more amenable to some of the points discussed below.

[15] Machamer et al. (2000, p. 8ff) and especially Craver (2007a, b, p. 114ff) discuss the chemical synapse example.

[16] Temporal sequence in biological m-mechanism may be less important, as they often exhibit more complex organization (Bechtel and Abrahamsen 2005; Abrahamsen and Bechtel 2012; Bechtel 2012). More generally, I include the more dynamically oriented definitions of mechanisms under the m-mechanism rubric.

Glennan's Law Glennan 1996, 2008, 2017; Craver 2001; Illari and Williamson 2012). This view applies to the m-mechanisms underlying cognition just as it does to the m-mechanisms underlying other natural phenomena (Bechtel and Abrahamsen 2005; Craver 2007b; Piccinini 2010; Kaplan 2011; Kaplan and Craver 2011; Piccinini and Craver 2011; Kaplan 2015; Glennan 2017).[17]

There are physical systems that do not satisfy the definition of a m-mechanism but that do satisfy the definition of a dynamical system. Take, for example, a rock (an example of Wimsatt's mere aggregates Wimsatt 1997). A rock has parts, such as the individual molecules, that are spatiotemporally characterized. These molecules might even engage in activity-like doings such as pinging off one another or arranging a lattice. Left at this, however, rocks are not yet responsible for some phenomenon and so are not yet m-mechanisms. Rocks are dynamical systems, however. In addition to the example of the rock, some systems do not engage in activity-like doings, such as Darden's broken clock (Darden 2006, p. 280ff) or static systems like a pillar supporting a roof (though see Illari and Williamson 2012, p. 130). Take for another example gravitational systems such as the attraction of two molecules. This system has spatiotemporally described parts and organization, but does not obviously engage in activity-like doings. So gravitational systems are dynamical systems but not m-mechanisms. Other systems appear to be constituted by activity-like doings but not parts, such as psychological systems underlying recognition or perception. A range of physical systems appear to lack an organization in virtue of being essentially constituted by random activity, such as gases or perhaps swarms, or featuring random organization, such as randomly connected neural networks.[18] Finally, other systems defy spatiotemporal characterization for a range of reasons, such as mathematical objects, institutions, or economies, and yet many are dynamical systems.

3.3 Productivity

Mere reference to a type of system such as a mechanism will not suffice for an explanation of some phenomenon. An explanation may refer to an m-mechanism but that does not entail that the explanation is a mechanistic explanation. Suppose one body attracts another with some gravitational force, and this force is explained by citing the universal law of gravitation. One or both of these bodies may be m-mechanisms, but this explanation is not mechanistic because the m-mechanisms did not produce the explanandum. In addition to the presence of a constitutive, m-mechanistic system, mechanistic explanations are ones that

[17] While many of the definitions cited are older, this physically focused view of mechanisms is still prevalent. Glennan, for example, says that "[e]ntities and activities are not abstract; they are fully determinate particulars located somewhere in space and time; they are part of the causal structure of the world. Sometimes there are abstract structures that can be characterized with mechanistic metaphors—but they are not mechanisms" (Glennan 2017, p. 20).

[18] Other, similar violations on the conditions of mechanisms are discussed by Levy and Bechtel (2016), who advocate expanding the concept of a mechanism to include problematic borderline cases.

feature systems that produce the explanandum (Miłkowski 2016). A system produces some phenomenon when the parts play roles in making a difference to the phenomenon.

Systems such as m-mechanisms must play some difference making role in explanations (Woodward 2003; cf. Miłkowski 2016; Klein 2017). A difference making role of an explanation is one such that if the difference maker were to change, then the explanandum phenomenon would change as well. Difference making further implies a relevance constraint for the system. *Ceteris paribus*, a change in a system that does not change the explanandum is not relevant to the explanation.

Another necessary condition to produce a phenomenon in a mechanistic explanation requires that the parts of the system play roles. A system could explain some phenomenon solely in virtue of the properties of and relations between the parts. For example, these relations could have mathematical properties that account for the phenomenon, such as the role of basis functions in explaining the representational power of neural networks (cf. Huneman 2018). These sorts of properties and relations are not the result of what the parts do in the system. A productive explanation, in contrast, features parts that do things. In sum, to produce the explanandum, the parts of a system must play roles that make a difference to the explanandum.

3.4 Conclusion

Explanations that feature systems are not *eo ipso* mechanistic explanations. To demonstrate that such explanations are mechanistic explanations, the system needs to satisfy the three conditions: constitution, mechanism, and productivity. Explanations of a phenomenon that feature constitutive, productive systems are merely potentially mechanistic explanations. Those explanations that feature constitutive, productive systems that do not satisfy the definition of an m-mechanism are pseudomechanistic explanations. I argue below that neurodynamical explanations are pseudomechanistic explanations.

4 Explanatory role of neurodynamical systems

What is the relationship between the neurodynamical explanatory schema and mechanistic explanation? In this section, I will consider and discard two takes on neurodynamical explanations. The first interpretation maintains that neurodynamical explanations are independent of constitutive explanations and mechanisms altogether. I will argue that this view is too radical, brushing aside important aspects of neurodynamical explanations. The second interpretation maintains that neurodynamical explanations are merely constitutive explanations. This latter constitutive interpretation captures only part of the neurodynamical explanatory schema and mischaracterizes how such explanations are constitutive. Other interpretations are required.

4.1 Neurodynamical systems and mechanisms are independent

On the first interpretation, neurodynamical explanations are not constitutive explanations and the neurodynamical systems that are featured in those explanations are not mechanisms. An example of this view is provided by Chemero and Silberstein's classification of explanatory patterns, who frame the question as "whether cognition is best explained mechanistically or dynamically" (Chemero and Silberstein 2008, p. 7). The independent interpretation rests on two claims. First, neurodynamical explanations feature explanatory models that "…allow one to abstract away from causal mechanical and aggregate micro-details to predict the qualitative behavior of a class of similar systems" (Chemero and Silberstein 2008, p. 12). Since these details are left out, the entities and activities of m-mechanisms are absent and hence dynamical systems are not mechanisms. Second, scientists who use dynamical systems theory to explain cognition are "employing differential equations as their primary explanatory tool" (Chemero and Silberstein 2008, p. 11), not m-mechanisms. The equations explain the phenomenon without appealing to constitution.

The independent interpretation seems to go too far, however. Neurodynamical explanations are constitutive explanations of a cognitive phenomenon by appeal to productive, organized sets of neurodynamical properties. The explanandum phenomenon is a cognitive capacity. Neurodynamical explanations explain such capacities by showing how neurodynamical systems perform functions for cognitive systems and in so doing have subcapacities such that when taken together, sets of such systems constitute the cognitive capacity being explained. This constitutive aspect is lost on the independent interpretation.

4.2 Neurodynamical systems non-mechanistically constitute cognitive phenomena

I now consider the constitutive interpretation, that neurodynamical explanations are constitutive explanations unrelated to systems with parts. The constitutive interpretation asserts that neurodynamical systems constitute and so are responsible for those cognitive phenomena but in a fashion irrespective of the causal-mechanical description of the world. This interpretation follows from a replacement of the second claim of the independent interpretation with a constitutive claim. Explanations of cognitive phenomena that invoke dynamical systems derive their explanatory power from constituting the explanandum phenomenon. But besides constituting phenomena by situating them in the world's causal-mechanical description, how else could a constitutive explanation work?

The deductive-nomological approach to explanation could be adapted to provide a type of constitutive explanation. Deductive-nomological approaches explain a phenomenon by showing that it can be rationally expected. For example, an explanandum phenomenon may be entailed by the specification of initial conditions and covering laws (Hempel and Oppenheim 1948). The deductive-nomological approach is often adopted in discussions of the explanatory power of dynamical systems (see,

e.g., Walmsley 2008 or Kaplan 2011). Though the deductive-nomological approach has been applied to constitutive explanation by deriving one theory from another theory plus bridge laws relating the two theories (Craver 2007b), the proposal here is to apply the initial conditions and laws framework to constitution.

Neurodynamical systems have subcapacities that correspond to the transformation of input signals into output signals. For example, the integrate-to-bound system takes motion evidence as input, integrates it over time, and outputs a summary of the evidence. These transformations are described mathematically; the integrate-to-bound system's evidence integration function has been described using a range of mathematical operations, including integration sensu calculus as described above. In mathematics, functions can compose: a function f: $x \rightarrow y$ and a function g: $y \rightarrow z$ can compose and result in a function $g \circ f$: $x \rightarrow z$. The constitutive interpretation analyzes constitution of cognitive capacities by the functional composition of the subcapacities of neurodynamical systems. Since function composition is ordered—not every function composition is commutative (it is not the case that $\forall f,g\ (f \circ g = g \circ f)$)—this notion of constitution is ordered as well. This implies that the ordered exercise of the subcapacities in the right sequence constitute the exercise of the capacity. A series of signal transformations that correspond to a sequence of mathematical operations functionally compose and deductively entail the output given some input. The series of function executions corresponds to the set of laws, and the initial conditions correspond to some input into the system.

The problem with this constitutive explanatory approach is that it provides an incomplete account of the neurodynamical explanatory schema. Neurodynamical explanations are explanations of cognitive capacities in terms of neurodynamical systems, organized collections of neurodynamical properties. The organization reflects both the dynamics of the systems as well as the way that different such systems can be put together. While the functions of neurodynamical systems undoubtedly do compose, the dynamics of these systems also compose, constituting the dynamics of the neurodynamical system. Constitution as function composition also does not provide an explanatory role for dynamical systems as part of the world's causal-mechanical structure. That latter aspect of neurodynamical explanations is absent from the constitutive interpretation. Of course, for the advocate of the deductive-nomological view, this concern is of no matter. Nonetheless, there may be ways of situating such systems in the causal-mechanical structure of the world.

5 Dynamical systems theoretic models as models of neurodynamical systems

Instead of further exploring the thesis that neurodynamical explanations are not mechanistic, I turn to discuss mechanistic interpretations of neurodynamical explanations. An intuitively plausible interpretation maintains that the schema describes how m-mechanisms help explain cognitive phenomena. The definition and defense of m-mechanisms in the literature discussed above suggests a partial interpretation of neurodynamical explanations. The dynamical systems theoretic (DST) models in those explanations are assumed to refer to m-mechanisms. However, these

descriptions can leave out spatiotemporal details of the putative entities and activities of m-mechanisms. Hence, DST models are incomplete models of m-mechanisms. Further, the absence of spatiotemporal details of the m-mechanism that instantiates the dynamics results in an incomplete mechanistic explanation relative to a description that includes those details (cf. Shagrir and Bechtel 2017). As Craver and Kaplan (2018) argue, the claim is not that merely providing more details makes a better explanation (the "More-Details-Better" thesis). Rather, it is that providing more relevant details makes a better explanation (the "More-Relevant-Details-Better" thesis), where the relevant details here are details about the m-mechanism that instantiates the dynamical system. On the partial interpretation, filling in the neurodynamical explanatory schema above results in a relatively incomplete explanation that leaves out relevant m-mechanistic details.[19]

This partial interpretation supposes that the targets of DST models are m-mechanisms. This supposition clashes with views that identify other targets besides m-mechanisms, such as processes construed as distinct from any type of mechanism (Dupré 2013), mechanisms in some sense other than m-mechanism [such as abstract mechanisms (Boone and Piccinini 2016) or structural mechanisms (Kuhlmann 2014; Felline 2018)], or as a distinct category of existent along the lines of the account of dynamical system outlined above (cf. Egan 2017; Woodward 2017).[20] Assuming the targets are m-mechanisms is a substantive thesis that requires defense.

In support of the partial interpretation, I will discuss three arguments inspired by positions that mechanists have taken. Proponents of m-mechanisms have crafted their view with an eye toward arguing against the independent interpretation above (see, e.g., Craver 2007b; Kaplan 2011; Kaplan and Craver 2011).[21] These arguments could be adapted for the current question of whether or not neurodynamical explanations are incomplete mechanistic explanations. All three arguments assume DST models target m-mechanisms but leave out relevant m-mechanism details for various reasons and so are incomplete models and as a result incomplete explanations. Neurodynamical explanations are incomplete because they either fail to meet a mapping requirement underlying cognitive neuroscientific explanations (Kaplan 2011; Kaplan and Craver 2011), only provide details about the changes in underlying mechanistic entities (Kaplan 2015), or merely incompletely describe m-mechanisms (Boone and Piccinini 2016). In opposition to the claim that DST models target m-mechanisms, the dynamicist might contend that a dynamical system distinct from an instantiating m-mechanism is the target. This possibility indicates an undefended enthymeme in the three arguments, that DST models target m-mechanisms. I maintain that these models target neurodynamical systems instead, thereby rejecting the partial interpretation, and defend this claim against several objections.

[19] I will usually elide the relativity of completeness in the following.

[20] I include processes here as a recent alternative to mechanisms. A discussion of the relationship between processes in the Dupré sense and dynamical systems goes beyond the scope of this essay.

[21] Many thanks to a reviewer who pointed out that these arguments are originally aimed at the independent interpretation.

5.1 Three arguments for the partial interpretation

In describing how mechanisms explain cognitive phenomena, some mechanists appeal to mapping constraints on explanatory models in cognitive neuroscience. Kaplan and Craver have formalized this constraint in their model-to-mechanism-mapping (3M) requirement (Kaplan 2011; Kaplan and Craver 2011), which can inspire an argument for the partial interpretation. (3M) states that

> In successful explanatory models in cognitive and systems neuroscience (a) the variables in the model correspond to components, activities, properties, and organizational features of the target mechanism that produces, maintains, or underlies the phenomenon, and (b) the (perhaps mathematical) dependencies posited among these variables in the model correspond to the (perhaps quantifiable) causal relations among the components of the target mechanism. (Kaplan and Craver 2011, p. 611).

Mechanisms play a role in explanations in cognitive neuroscience by serving as the targets for DST models. Though not explicit in (3M), the relevant sense of mechanism is m-mechanism (see, e.g., Kaplan and Craver 2011, p. 605ff). The dynamics described by those models pick out m-mechanisms in virtue of the variables in the models denoting entities, activities, and so forth of m-mechanisms and the transformations in the models denoting causal relations between entities of those m-mechanisms. However, these models leave out much relevant details about entities and activities. Hence, neurodynamical explanations are incomplete explanations. Call this the mapping argument.[22]

The mapping argument assumes that the targets of DST models are m-mechanisms. However, as discussed above, models of systems are distinct from the systems themselves, and the targets of the models may not be m-mechanisms. For example, in the case of the integrate-to-bound system in LIP, the targeted system could correspond to a neurodynamical system constituted by single (or multi-) neuronal dynamics or to a m-mechanism constituted by the spatiotemporal entities underlying those dynamics (or both). The model of this integrate-to-bound system, which specifies a set of variables and their transformations, is distinct from these dynamics. Both the neurodynamical system and the instantiating m-mechanism are prima facie possible targets for the model.

On the mechanics argument for the partial interpretation, dynamics are construed as descriptions of the temporal evolution of entities of m-mechanisms (Piccinini and Craver 2011; Kaplan 2015).[23] Kaplan contends that "dynamical models are… well suited to reveal the temporal organization of activity" in neural m-mechanisms

[22] I don't mean to imply that Craver, Piccinini, Kaplan and other mechanists would endorse the mapping argument. Rather, I am taking their stated positions on the proper role of dynamical systems models in cognitive neuroscience explanations as one way of arguing for the partial interpretation, which may or may not be a use of such mappings which these philosophers would endorse.

[23] All three arguments are compatible with each other and I don't mean to suggest that philosophers who endorse one could not also endorse another.

(Kaplan 2015, pp. 759–760). On this inspiration for the partial interpretation, DST models are models of the operation of entities of m-mechanisms, descriptions of the temporal (and possibly spatial) dynamics of m-mechanism components (Piccinini and Craver 2011; Kaplan 2015).[24] Once again, since these models leave out much relevant m-mechanistic detail, neurodynamical explanations are incomplete. The mechanics argument also assumes that the targets of DST models are m-mechanisms. If the target of the modeling effort is the dynamics of m-mechanisms, then those models may indeed merely reflect the instantiating m-mechanisms. But the target of the modeling effort may equally well be the neurodynamical systems themselves independent of the instantiating m-mechanism. On the former target, relevant missing details include spatiotemporal details about entities and activities; on the latter, these spatiotemporal details may be irrelevant. Consequently, the mechanics argument fails to demonstrate that neurodynamical explanations are incomplete mechanistic explanations.

On the incomplete description argument for the partial interpretation, DST models are incomplete descriptions of mechanisms, whether schemata, descriptions of mechanisms with details deliberately omitted, or sketches, descriptions of mechanisms that leave out unknown details. A number of extant statements about sketches and schemata can lend inspiration to arguments for the partial interpretation. Piccinini and Craver have argued that DST models provide a functional description of an m-mechanism, where "functional descriptions are elliptical mechanistic descriptions", specifically sketches of m-mechanisms that leave out unknown spatiotemporal and other details (Piccinini and Craver 2011, p. 307).[25] But an alternative prima facie viable position is that these are models that target neurodynamical systems and so don't need to include details specific to the instantiating m-mechanisms.[26] According to Boone and Piccinini, some details are left out because "[i]dentifying and explaining those different mechanism types requires omitting the idiosyncratic

[24] Cf: "Organization is… a necessary part of most moderately complex mechanisms such that perturbing either the spatial organization or temporal dynamics of a mechanism, even while the components and their activities remain unchanged, can have appreciable (even catastrophic) effects on its performance. Thinking about mechanistic explanation, then, it is clearly insufficient to describe only the properties and activities of the component parts in a given mechanism without giving adequate weight or attention to the spatial and/or temporal organization of those parts and activities. Often this point is underappreciated or lost when considering the nature of mechanistic explanation…. understanding the dynamical "structure" of a mechanism can be just as important as understanding its physical structure" (Kaplan 2015, pp. 774–775).

[25] They go on to assert that "[t]he idea that functional description is somehow autonomous from details about mechanisms involves a fundamental misunderstanding of the nature of functional attribution in sciences like cognitive neuroscience" (Piccinini and Craver 2011, p. 307). This overlooks at least one clear alternative, that cognitive neuroscientists are concerned with functional descriptions sensu dynamics, and neuroscientists *simpliciter* are concerned with m-mechanisms. A helpful analogy here is between a car designer or engineer and a car builder or mechanic. The designer or engineer might only care about the functional descriptions of the parts, leaving it to the builder or mechanic to determine the appropriate m-mechanisms. Something similar could be said about cognitive neuroscience and neuroscience. On such an analysis, there is some sense in which functional description is autonomous from m-mechanisms. I will forego further discussion of the issue of autonomy for another time, but see the very nice discussions in Kaplan (2017).

[26] See below in Sect. 6 for an extended discussion of functional analysis.

details of less abstract types of mechanism in order to reach a description that is general enough to denote the relevant features that the less abstract types have in common" (Boone and Piccinini 2016, p. 693). Other details are left out in order to capture regularities across m-mechanisms by "…isolating features that are shared by mechanisms that occur within radically different systems and may even occur at different levels of organization…" (Boone and Piccinini 2016, p. 694). Dynamical systems are descriptions either of types of m-mechanisms or of regularities across many different particular m-mechanisms and so should be considered mechanism schemata. But this assumes these are models of the instantiating m-mechanisms as opposed to models of neurodynamical systems in their own right, an assumption that requires independent justification (cf. Egan 2017). In sum, the incomplete description argument too assumes the targets of these DST models are m-mechanisms, and so requires further defense to demonstrate the partial interpretation.

5.2 Objections

I will now consider some objections to the claim that the DST models present in neurodynamical explanations can target dynamical systems distinct from their instantiating m-mechanisms. First, the mechanist might object that for the dynamics targeted by models to exist, they must be dynamics of m-mechanisms and so the targets of the models are m-mechanisms. The definition of a dynamical system given above explicitly states that these systems are token identical to subsets of the dynamics of their instantiating physical systems. Plausibly some of the physical systems instantiating the dynamics are m-mechanisms, as many philosophers (e.g., Craver 2007b) have argued for the brain. Consider for example Bechtel and Abrahamsen's take on mechanistic dynamics (Bechtel and Abrahamsen 2010). They claim that dynamical systems theoretic models are "…used to better understand the functioning of a mechanism whose parts, operations, and organization already have been independently determined" (Bechtel and Abrahamsen 2010, p. 322). They note that modelers "probe how the mechanism's organization, parts, and operations are orchestrated in real time to produce dynamic phenomena…" (Bechtel and Abrahamsen 2010, p. 322). On this objection, the dynamics are spatiotemporal details of entities and activities of the m-mechanism and so the models do target m-mechanisms.

In reply, I contend that m-mechanisms, non-m-mechanism dynamical systems, and models are distinct. The various categories divide up systems in different ways according to identification, re-identification, and classification conditions (in short, sortal conditions; cf. Strawson 1959 or Quine 1960). Definitions provide sortal conditions. The discussion above defined m-mechanisms as those spatiotemporal systems that are sets of entities and activities organized so as to be responsible for the production of some phenomenon. The sortal conditions for m-mechanisms then include the entities, activities, organization, and productive roles in systems. Dynamical systems herein are defined as the objects, properties, and relations of some system and the changes in those objects, properties, and relations. While neurodynamical systems happen at a place (or in a volume) and over time, their sortal conditions need not specify those properties. As a result, m-mechanistic sortal conditions may

be different from the sortal conditions for dynamical systems. Granted these distinct sets of conditions, m-mechanisms can be distinct from dynamical systems, and so models can target one type of existent without targeting another.

A second related objection denies that there are distinct explanatory roles for these different categories.[27] Dynamical systems are token identical to the dynamics of physical systems. If neural systems are m-mechanisms, then neurodynamical systems are token identical to the dynamics of m-mechanisms. So, the objection goes, there is no explanatory role for a neurodynamical system that is distinct from the explanatory role of its instantiating neural mechanism.

This objection raises deep issues related to identity and explanation. In defense against this objection, I reject that token identity of neural m-mechanisms and neurodynamical systems entails identical explanatory roles. First, if two systems have identical explanatory roles then they have identical counterfactual explanatory constraints. Second, token identical neural m-mechanisms and neurodynamical systems have distinct sortal conditions. Third, distinct sortal conditions entail distinct counterfactual constraints. So, neural m-mechanisms have different counterfactual explanatory constraints than neurodynamical systems. By modus tollens, they have different explanatory roles.

Three considerations can be provided in favor of the first premise, that identical explanatory roles imply identical counterfactual constraints (for a general argument that model-based explanation centrally involves situating a phenomenon in a pattern of counterfactual dependence, see Bokulich 2011). Each of these considerations is not airtight, but they illustrate how explanatory roles can imply counterfactual constraints. First, explanations imply counterfactual failure constraints on explanatory role. Something plays an explanatory role such that if the physical set-up were to change and as a result the phenomenon being explained fails to occur, then whatever plays that role must explain that failure. Take, for example, the explanation of lighting a match. After a match is struck, the match alights due to phosphoric ignition and fuel in the form of oxygen and sulfur from the match head. This in turn ignites the wood. Now suppose the match is damp and fails to light when struck. If the entities in the match head play an explanatory role in explaining why the match lights, then those entities must change when the match is damp. Second, explanations imply counterfactual modulation constraints, where some things may play an explanatory role such that if the physical set-up were to change and as a result the explanandum phenomenon changes, then whatever plays that role must explain that change. Suppose striking the match at an acute angle will more slowly light the match. This is explained by the fact that it is a match but not by the fact that it is a piece of wood. Third, explanations imply counterfactual persistence constraints, where some things may play an explanatory role such that if the physical set-up were to change and the explanandum phenomenon does not change, then whatever plays that role must persevere through the change in the physical set-up. Changing from oak to pine does not change the fact that the match ignites. This persistence of the explanandum is explained by the fact that it is a match and not a piece of oak.

[27] I thank an anonymous reviewer for this objection.

The other premises are more straightforward. The second premise follows from the analysis of neurodynamical systems and the definition of m-mechanism above. The third premise follows from the definition of sortals. By negation of the consequent of the first premise, different counterfactual constraints imply different explanatory roles. By modus tollens, neurodynamical systems and neural m-mechanisms have different explanatory roles even if they are token identical. Granted the neurodynamical explanatory schema presented above, neurodynamical systems and neural m-mechanisms have distinct explanatory roles in explaining cognitive phenomena.

An objection charges this argument with an equivocation. On this objection, the counterfactual constraints cited in the first premise that arise from explanatory roles are different from those cited in the third premise that arise from sortal conditions. Because the counterfactual sortal constraints are different from the counterfactual explanatory constraints, a difference in the sortal constraints does not imply a difference in the explanatory ones. For example, a system may change without changes in the explanandum phenomenon. The neurons that constitute the m-mechanism in LIP could change and yet the counterfactual explanatory constraints determined by the explanatory roles for the mechanism may not dictate a corresponding change in the explanandum phenomenon. Perhaps other neurons are able to function the same as the original ones for the integration of motion evidence. Though such a change in entity would result in a changed m-mechanism, the explanandum may remain the same because different neurons can do the job equally well. Nonetheless, the explanandum phenomenon, the integration of evidence, remains the same. So, explanatory counterfactual constraints are not the same as sortal counterfactual constraints. The objector concludes that the counterfactual explanatory constraints are distinct from the sortal ones, and one can't infer from distinct counterfactual sortal constraints to distinct explanatory roles.

In reply, the counterfactual constraints imposed by sortals are in some cases identical to those counterfactual constraints imposed by explanatory roles. Insofar as they are identical, the argument will go through. This connects sortal conditions for types and explanatory roles in a way that refutes the thesis that token identity is sufficient for identical explanatory roles. In these cases, explanatory roles and types have identical counterfactual explanatory constraints. Even if instances of two types of entity are token identical, the entity qua one type may have a different explanatory role than the entity qua the other type given the differences in the counterfactual explanatory constraints described by their sortal conditions. What remains to be shown, then, is that neurodynamical explanations feature counterfactual explanatory constraints that line up with the counterfactual sortal constraints for neurodynamical systems.

Suppose some neural m-mechanism and its changes are token identical to some neurodynamical system that plays some explanatory role. Now consider counterfactual persistence constraints. Elements of the neural m-mechanism can be removed, thereby changing the type identity of the m-mechanism. This change in the mechanism does not entail a change in the neurodynamical system, however, because the dynamics may remain the same. The neurodynamical system may remain the same because the substrate for the system are not the changes in the m-mechanism that give rise to the activity level of neurons but rather the activity level itself and there

are many different ways to generate the same level of activity. In particular, suppose that despite the changes in the neurons, there are always integrate-to-bound dynamics from trial to trial. For example, some single neurons in LIP exhibit integrate-to-bound dynamics, whereas others do not—but if the neurons that do not exhibit such dynamics are taken as a group, then those dynamics are present for the population (Meister et al. 2013). These dynamics may be sufficient for motion integration. The system could sometimes utilize single neuron dynamics and sometimes not, and even if the single neurons are silenced, the population may still exhibit the dynamics needed to integrate motion information. So, some changes in the m-mechanism do not entail a change in the explanandum phenomenon. However, if the integrate-to-bound dynamics are absent, then the cognitive system does not sum up the evidence as to motion direction. In this latter instance, though, the identity of the m-mechanism can be the same because a specific dynamic profile may not be part of the sortal conditions on the mechanism. This implies that the m-mechanism and dynamical system have different counterfactual sortal constraints in virtue of which they play different explanatory roles.[28] Furthermore, the objection that the neurodynamics that play the explanatory role are always a m-mechanism because they are token identical to part of a m-mechanism is to commit a mereological fallacy. A part of the neural m-mechanism such as the neurodynamics does not imply that the neurodynamics are also a m-mechanism.

Austin has critiqued this line of reasoning as follows. After defining a mechanism in terms of the temporal and spatial organization of its entities and activities (Austin 2016, p. 644), Austin argues that "…it's open… to the defender of a mechanistic ontology to… reform her concept of 'mechanism'… in such a way that the persistence of a particular mechanism need not depend… upon… a static set of entities performing a static set of activities" (Austin 2016, p. 654). Entities and activities can change without the identity of the mechanism thereby changing.

Two points in reply. First, some change is surely permitted. But there must be constraints as well, lest every arbitrary collection of entities and activities count as the same m-mechanism. However, the amount of variation permitted may be less than that which may be found underlying neurodynamical systems. According to Austin, "mechanisms are ontologically defined/individuated by their unique fourfold structure—that is, their (1) type and number of entities and (2) their spatial organisation, and their (3) type and number of connective activities and (4) their spatial and temporal organisation. Thus, what it is to be a particular mechanism is to be an instance of a specific fourfold structure" (Austin 2016, p. 644). How much variability in this fourfold scheme is permitted is unclear. However, I take it that a complete replacement of any of the four elements of the mechanistic structure is sufficient to yield a new m-mechanism. This is supported by Austin's claim that "distinct permutations of the values of (1)–(4) constitute distinct mechanisms, and so in any particular case the alteration of any of those values amounts to the effective dissolution of that mechanism" (Austin 2016, p. 644), where a wholesale replacement of one of

[28] Ceteris paribus, of course.

the elements of the structure is sufficient for a distinct permutation and so a different mechanism.

Are there distinct permutations of one of these elements while retaining the same dynamics? The Meister et al. study suggests that if one were to silence all the single neurons that exhibited integrate-to-bound dynamics, then the system might still rely on the rest of the population for those dynamics. The other members of the population in concert exhibit the required changes for integrating evidence. So they could serve in place of the silenced neurons. But ex hypothesi, these are exclusive sets, and so the mechanism would be a token of a different type. Of course, this is speculative and the normal functioning of LIP might utilize both types of cells, which implies that silencing only one type might not count as a wholesale replacement.

Second, the presence of variability *simpliciter* in the definition of an m-mechanism is not entirely to the point. Rather, the ways that an m-mechanism can vary fail to line up with the ways that a neurodynamical system can vary, and the ways that m-mechanisms do vary does not coincide with changes in the explanandum phenomenon, whereas the ways that the neurodynamical systems vary do so coincide. Specifically, consider the case of counterfactual failure constraints. For example, the hallmark ramp up in LIP firing rates is markedly decreased during error trials compared to correct trials in LIP (Roitman and Shadlen 2002). Absent the integrative trajectory toward the threshold, the neurodynamical system malfunctions, evidence is not integrated, and the cognitive system commits an error. Because these dynamics are not included in the mechanistic sortal conditions, the m-mechanism does not vary in these situations. And yet, the explanandum phenomenon does. Intuitively, the failure mode for the m-mechanism does not match the failure mode for the function of integrating evidence, whereas the failure mode for the neurodynamical system does. This implies that because the m-mechanism counterfactual sortal constraints fail to match the counterfactual explanatory constraints, m-mechanisms are not the relevant type under which falls a particular token system that explains some cognitive phenomenon.

5.3 Causal-mechanical structure and the partial interpretation

Another argument for the partial interpretation proposes that neurodynamical systems causally depend on the properties of entities of m-mechanisms. Because the dynamics change only as a function of changes in the m-mechanism's entities or activities, any explanation that cites the dynamics is incomplete without specifying how the m-mechanism's entities or activities change to produce the dynamics. In particular, the causal relations between entities are crucial to these explanations (cf. Craver and Kaplan 2018, p. 11ff). The counterfactual relations discussed above require an underlying causal structure that can only be provided by an m-mechanistic explanation that situates the phenomenon in the causal structure of the world. So, DST models are incomplete: they leave out the relevant causal relations needed for mechanistic explanations.

Let's grant the mechanist that the causal powers lie in the entities and activities of m-mechanisms. That still doesn't entail that neurodynamical explanations are

incomplete without the m-mechanistic details. But, the mechanist replies, then the neurodynamical system remains unexplained. In response, the supporter of neurodynamical explanations could happily agree with the mechanist that the neurodynamical system remains unexplained, for that was never the target of the explanation anyway. The explanandum is the cognitive capacity, and that is explained by completing the neurodynamical explanatory schema.[29] But that explanation does not entail that completing the schema yields an incomplete explanation of the cognitive phenomenon for lack of leaving out m-mechanistic detail. In rejoinder, the objector could insist that if neurodynamical systems explain cognitive phenomena, then in the absence of such causal detail the cognitive phenomenon fails to be situated in the causal-mechanical structure of the world. This objection is, at heart, a demand for an account of the explanatory power of neurodynamical systems. I postpone my reply to this demand to the discussion below that argues in favor of the pseudomechanistic interpretation.

The point of the foregoing was to evaluate the partial interpretation that the neurodynamical explanatory schema outlines a form of incomplete mechanistic explanation. The partial interpretation contends that neurodynamical explanations are incomplete mechanistic explanations because DST models are incomplete models of m-mechanisms. The three arguments for the partial interpretation overlook the possibility that neurodynamical systems could be the target of DST models to which the schema appeals. If those neurodynamical systems are the targets, though, then neurodynamical explanations need not fill in the m-mechanistic details. This leads to one last even stronger interpretation, the pseudomechanistic interpretation.

6 Neurodynamical systems and the pseudomechanistic interpretation

On the pseudomechanistic interpretation of the neurodynamical explanatory schema, neurodynamical explanations satisfy the constitutive and productivity conditions on mechanistic explanations but fail to satisfy the mechanism condition because neurodynamical systems are not m-mechanisms. The pseudomechanistic interpretation agrees with the partial interpretation that neurodynamical explanations are incomplete mechanistic explanations. However, unlike the partial interpretation, the pseudomechanistic interpretation does not entail that neurodynamical explanations are incomplete because dynamical systems theoretic (DST) models target m-mechanisms. The pseudomechanistic interpretation also identifies different systems—and so different parts—than the partial interpretation to fill the constitutive and productive roles of a mechanistic explanation. In reply and as an attempt to fully incorporate neurodynamical explanations as mechanistic explanations, some mechanists are willing to relax the requirements on m-mechanisms in a way that seems to subsume dynamical systems (Kaplan and Craver 2011; Piccinini and Craver 2011). I argue

[29] Recall that I am simplifying the true complexity of explanations of cognitive phenomena, which require many dynamical systems executing many functions.

that no light revision of the definition of a m-mechanism will suffice. Other philosophers argue that these explanations are to be understood as a species of functional analysis and assimilate functional analysis to m-mechanisms. I argue that while neurodynamical explanations do entail functional analysis, the type of functional analysis is both novel and consistent with pseudomechanistic explanation. Recall above that the m-mechanist might challenge the constitutive explanatory power of the neurodynamical explanatory schema. In reply, I briefly outline how pseudomechanistic neurodynamical explanations fit into the causal-mechanical framework. In addition, pseudomechanistic explanation, like mechanistic explanation, requires systems to produce their explananda. I also outline the case for neurodynamical systems producing cognitive phenomena. In conclusion, neurodynamical systems are either a distinct ontological category from mechanisms or mechanisms in some other sense than m-mechanism (Kuhlmann 2014; Boone and Piccinini 2016; Felline 2018), and neurodynamical explanations are at best pseudomechanistic explanations. I end with a brief coda on the importance of neurodynamics for cognition.

6.1 Could neurodynamical systems be mechanisms?

I contend that neurodynamical explanations satisfy the constitutive constraints on mechanistic explanation. This implies that neurodynamical explanations contain organized systems with parts that make up the explanandum. In order to make up the explanandum, the parts must have functions or play roles in producing it.

What are the parts? In a neurodynamical explanation of a cognitive phenomenon, the parts are the neurodynamical systems themselves. For example, the integrate-to-bound system is a part, and its function is to integrate evidence for the cognitive system. The explanation of a cognitive phenomenon requires completion of the neurodynamical explanatory schema for each subcapacity into which the cognitive phenomenon was analyzed. Each subcapacity is performed by a neurodynamical system, a part of the cognitive system. Hence, neurodynamical explanations feature parts playing roles.

Second, the neurodynamical systems themselves may have parts. Recall that dynamical systems are collections of objects, properties, and relations and the changes in them. The substrate makes up the state space for the system, the collection of possible determinate objects, properties, and relations. (For convenience, I will call a point in the state space a property—think of it as a conjunction of all such maximally specified triples of objects, properties, and relations.) These state spaces have features that could be considered the parts of dynamical systems. Dynamical systems are often sorted into types according to qualitative features of their state space such as attractors, limit cycles, bifurcations and the like. These are structures seen in the changes in the substrate such as the threshold attractor in the integrate-to-bound system.

Consider now that the parts of neurodynamical systems are those features of dynamical systems. These properties can have functions and play a role in the production of phenomena. Call the functions and roles the performances of the parts of the dynamical system. I will consider cognitive performances only: functions or

productive roles that are specified in cognitive terms and that are performed for the cognitive system.[30] What are the performances for such parts of the system? The performances will depend on the context in which the dynamical system is used. For example, the function of the threshold may be to signal or represent the crossing of a decision boundary,[31] to start the reset of the system, or to initiate motor responses. Insofar as the threshold performs some function for the system, the threshold helps produce the explanandum phenomenon. The threshold is a productive, role-filling part of the integrate-to-bound system.

The takeaway, then, is that the supporter of neurodynamical explanations as explanations by systems with parts has several options available for the parts. Neurodynamical systems themselves help explain some cognitive capacity in virtue of being parts that perform functions for the cognitive system for that capacity. These neurodynamical systems may themselves have parts as well, the grosser-grained dynamical features of those systems. The best characterization and evaluation of these parts depends on which aspects of neurodynamical systems are explanatory.

6.2 Are neurodynamical system m-mechanisms?

Grant then that these neurodynamical systems are systems with parts that can help produce the explanandum and, so, neurodynamical explanations are constitutive. Can the dynamical systems cited in neurodynamical explanations be considered m-mechanisms? Recall that an m-mechanism is an organized set of entities characterized by their physical properties (such as spatiotemporal location, shape, electrochemistry, etc.) and activities characterized by their physical properties (such as rate, duration, material changes, etc.) that causally produce and are responsible for some phenomenon. To accommodate neurodynamical systems, the analysis of m-mechanisms provided above would have to be augmented. In consideration of the role of mechanisms and dynamical systems in cognition, Kaplan and Craver reassessed what makes a m-mechanism and noted that the entities "…need not be spatially localized within the system. Nor need their activities be sequential, from beginning to end…" (Kaplan and Craver 2011, pp. 605–606). But Kaplan and Craver do not go far enough in loosening the constraints on m-mechanisms. The problematic elements of m-mechanisms above are still present: spatiotemporal characterization of entities and activities. Descriptions of parts and their productive roles in neurodynamical systems need not include these details. Piccinini and Craver similarly suggest that the entities of m-mechanisms need not be "…neatly spatially localizable, have only one function, are stable and unchanging, or lack complex or dynamic feedback relations with other components. Indeed, a structural component might be so distributed and diffuse as to defy tidy structural description, though it no doubt has one if we had the time, knowledge, and patience to formulate it" (Piccinini and Craver 2011, p. 291). Once again, the suggested revision is in the right direction but

[30] There may also be non-cognitive performances performed for the system.

[31] In order to avoid debates over representation, I am deliberately imprecise about whether these parts are representations.

not strong enough. Certain details of parts, such as their location, shape, or orientation, may be irrelevant to the description of the neurodynamical system. Further, the productive roles of parts of dynamical systems are not necessarily individuated by spatiotemporal location, rate, or duration like activities are. Nor do neurodynamical systems' structural or organizational properties get specified in such a concrete way.

The standard approach to m-mechanisms would need a deeper revision than the superficial ones considered so far. The definition of m-mechanisms focuses on the machine-like: entities are defined in terms of their spatiotemporal properties, activities are defined in terms of their spatiotemporal properties, and so forth. But cognitive systems are defined in terms of their functional properties in two senses, the way they behave and the subcapacities that they possess. First, their parts are described in terms of the structure of their behavior, such as the structure of the state space through which the system travels. The integrate-to-bound system is type individuated by the structure of its state space. Second, their parts are functionally defined in the way that, for example, a fuel injector is a part of a car: its capacities exhaust the description of the part, namely, to inject fuel, and the spatiotemporal and other properties are left out of the description. For the cognitive system during noisy perceptual decision making, the integrate-to-bound system is a part that integrates motion evidence to a bound for perceptual decision making. Similarly, the parts of a neurodynamical system can be described by their behavior or capacities such as an integrate-to-bound threshold initiating action selection and resetting the system. A minor widening of the definition of a m-mechanism does not accommodate neurodynamical systems, including too many details about entities and activities and too little about these behavioral and functional dynamics.

6.3 Functional analysis

The foregoing comments suggest that neurodynamical systems and neurodynamical explanations rely on functional analysis. An argument in favor of the partial interpretation and against the pseudomechanistic interpretation can be constructed on the grounds that functional analysis results in DST models that require m-mechanistic details for explanatory power. In their 2011 paper, Piccinini and Craver argue that functional analysis results in mechanism sketches.[32] m-mechanisms are made of entities (in their terms, 'components') that have functional and structural properties: "Components have both functional properties—their activities or manifestations of their causal powers, dispositions, or capacities—and structural properties—including their location, shape, orientation, and the organization of their sub-components" (Piccinini and Craver 2011, p. 291). In the case of structural components, "…functional analysis… is a promissory note on (a sketch of) a mechanistic explanation" (Piccinini and Craver 2011, p. 300). This follows from their description of structural properties as spatiotemporal properties and a view of functional analysis as saying what components do but now how they do it. In the case of functional components,

[32] I thank a reviewer for their request for a response to Piccinini and Craver's arguments regarding functional analysis.

components are "…functionally individuated… or black boxes" (Piccinini and Craver 2011, p. 300). By their lights, such functional individuation implies causal properties. Some such functional individuations can then be ruled out by discovering components that lack the implied causal properties. Piccinini and Craver first note that "[f]unctional analysis borrows its explanatory legitimacy from the idea that functional explanations… capture something of the causal structure of a system" (Piccinini and Craver 2011, p. 306). As a result, "[l]earning about components allows one to get the right functional decomposition by ruling out functional decompositions that are incompatible with the known structural details" (Piccinini and Craver 2011, p. 306). They conclude that "the search for mechanistic details is crucial to the process of sorting correct from incorrect functional explanations" (Piccinini and Craver 2011, pp. 306–307).[33] The DST models that result from the sort of functional analysis above require m-mechanistic details and, in their absence, are merely incomplete mechanism sketches.

On my view, neither sense of component entails incomplete DST models or m-mechanistic explanations. Take their first reading of components as structural. Structure could mean spatiotemporal structure, consistent with the definition of m-mechanism above. But structure could also mean the qualitative features present in dynamical systems' state spaces. Call that dynamical structure. Recall that a mechanism sketch is a description of an m-mechanism that leaves out unknown details. On the first sense of structure, their claim that functional analysis yields a mechanism sketch is accurate. But on the second sense, the claim does not follow. As I've just argued at length, DST models can target dynamical systems. These dynamical systems can possess dynamical structure whose elaboration does not require providing m-mechanistic details. Likewise, their claim about functional explanations capturing the structure of a system is similarly equivocal. The structure captured by functional analysis could be spatiotemporal structure as they implicitly assume or it could be dynamical structure.

Now take their second reading of components as functional, where this functional reading requires m-mechanistic details in order to sort correct from incorrect functional analyses. Key to neurodynamical explanations are two dimensions of analysis. The behavior of a system can be dynamically described, how the substrate changes, or the system can be described in terms of subcapacities, the functions the system performs for some other containing system. A successful neurodynamical explanation involves a mapping between the behavior and the subcapacities. Their account seemingly overlooks the normative grounds provided by this mapping. If one views

[33] Piccinini and Craver present a second argument as well. They claim that "…explanations that capture these mechanistic details are deeper than those that do not" (Piccinini and Craver 2011, p. 307) for two reasons: first, "…it allows one to expand the range of phenomena that the model can save" and second, "…knowledge of the underlying components and the structural constraints on their activities affords more opportunities for the restoration of function and the prevention of calamity or disease" (Piccinini and Craver 2011, p. 307). I do not have space to adequately address this argument. However, note that neurodynamical explanations can be seen to capture more phenomena than those captured by m-mechanisms in virtue of different m-mechanisms giving rise to the same neurodynamical system. Their pragmatic point is well-taken, though one can intervene on dynamics too.

functional analysis in purely capacity terms, then without parts to possess those capacities, the functional analysis is utterly ungrounded. But on the other hand, if one admits of behavioral descriptions of parts independent of their capacities, then one can once again ground the capacities in the system. The dynamical description of a subsystem's behavior can provide the details that allow sorting correct from incorrect functional explanations.

This distinction between the subcapacities and behaviors of systems allows for the description of systems while remaining agnostic about instantiating mechanisms (cf. Shapiro 2016). These descriptions yield dynamic or functional parts and so need not include spatiotemporal details that are relevant only to the m-mechanism. The functions of parts are often not specified in terms of the neurophysiological entities and activities that underlie this functioning, like the types of specific neurons, the flow of specific ions, or the action of neuromodulatory molecules. And the dynamics of the system such as changes in state, the trajectories through state space, and features of this state space need not specify the neurophysiological details required by m-mechanisms. Hence, neurodynamical explanations are not mechanistic sketches.

6.4 Constitution and productivity

I have argued that neurodynamical systems are not m-mechanisms and that neurodynamical systems are parts of cognitive systems and, further, may themselves contain parts. I maintain that neurodynamical systems constitute and produce cognitive phenomena. Neurodynamical explanations are thus pseudomechanistic explanations.

Mechanistic explanations, as discussed previously, gain their explanatory power by showing how m-mechanisms constitute the explanandum phenomenon. The constitution of the explanandum by the m-mechanism places the explanandum in the causal-mechanical structure of the world. The entities of the m-mechanism are causally active objects and the activities of those entities are some of the ways that they cause things. The description of how those entities and activities produce the explanandum is to describe how the explanandum fits in that causal-mechanical structure. Because neurodynamical explanations are merely pseudomechanistic explanations, they do not describe entities and activities that fit into this causal-mechanical structure in the same way that mechanistic explanations do. But then, whence derives their constitutive explanatory power? Why are neurodynamical explanations constitutive explanations?

To answer this question, the concept of causal-mechanical structure needs to be unpacked. Structural properties of the world emerge from the entities, activities, and their causal interactions. The static structural properties are properties and relations of the entities and activities that make up the causal-mechanical organization. But there are also dynamic structural properties, the changes that result from the entities engaging in activities and causal interactions. This structure can serve as an additional source of explanatory power. In particular, this structure can itself be described and explananda can be shown to be constituted by elements of this structure. One reason, then, that neurodynamical explanations possess explanatory power is that they situate cognitive phenomena in the structure of the world

by showing how those phenomena are constituted by the dynamics of the world's causal-mechanical structure. But as I mentioned above, these dynamical properties themselves change, and those changes can change, and so on. So neurodynamical explanations can also possess explanatory power because they situate cognitive phenomena in the world's structure by showing how the dynamics that result from changes in the world's dynamical structure constitute cognition. Constitutive dynamical explanation can result from the changes in the entities and activities of the mechanisms in the world, but it can also result from changes in the changes in … the changes in those mechanisms. I submit that neurodynamical properties are often these higher-order changes whose substrate are lower-order changes in yet further lower-down mechanical substrates. The diehard mechanist might object here that this is not causal-mechanical structure, but I don't think this objection carries force because the cognitive explananda are still situated in the properly elaborated causal-mechanical structure of the world.

Besides constitution, pseudomechanistic explanations require systems to produce their explananda. To produce the explanandum, a system must be a difference maker: if the system were to change, then the explanandum phenomenon does as well. The difference making condition just is the set of counterfactual constraints placed on systems that explain phenomena. Neurodynamical systems perform functions for cognitive systems such that if different inputs were received by the neurodynamical system, then the system would be behave differently. This is prima facie evidence that these systems are difference makers. For example, pulses of motion evidence result in specific changes in the integrate-to-bound system in LIP that are subsequently evident in animal's behavior (Huk and Shadlen 2005).[34]

In addition to being a difference maker, the parts of the system play roles that help produce the explanandum. In the case of neurodynamical explanations, the explanandum is a capacity of some cognitive system. The parts are the neurodynamical systems and the roles they play are the functions they perform for cognitive systems such as integrating motion evidence. As illustrated above, the integrate-to-bound system integrates evidence for perceptual decision making. This function is illustrated by failures to integrate evidence, where the integrate-to-bound dynamics are absent. That failure implicates the neurodynamical system in the production of the explanandum. In sum, neurodynamical explanations are constitutive, productive but non-m-mechanistic explanations, that is, they are pseudomechanistic explanations.

I argued above that neurodynamical systems themselves may have as parts the features of its dynamical structure. These parts of the integrate-to-bound system play roles in the system that help produce the subcapacity to integrate motion evidence, such as, for example, the threshold that may represent or indicate the culmination of

[34] New evidence questions the role of the integrate-to-bound system in LIP, as inactivating the region does not affect behavior (Katz et al. 2016). But there are other areas that exhibit these dynamics during the task (Ding and Gold 2011; Ding and Gold 2012; Ding and Gold 2013; Hanks et al. 2015; Brody and Hanks 2016), so this may simply suggest that the system is not actually instantiated in LIP or that the dynamical properties in LIP are a read-out of those properties elsewhere in the brain. Also, the nature of the explanatory enterprise can be revealed even if the specifics of the case study are false.

the accumulation of motion evidence or that initiates action selection. The capacities of the neurodynamical systems themselves may also be explained by instances of the neurodynamical explanatory schema. Taking both the explanations of the capacities of the cognitive system and the potential explanations of the subcapacities by the dynamical structure of these neurodynamical systems, a hierarchical explanation of cognitive phenomena emerges, with cognitive systems constituted by neurodynamical systems that are themselves constituted by dynamical structures in their state space.

In conclusion, I would like to stress that the issue of the mechanistic status of these dynamical systems should not occlude their scientific importance. Indeed, focusing too much debate on whether these systems are mechanisms, whether neurodynamical explanations require mechanisms in some sense and to what degree, and related questions can obscure the role that dynamical systems play in guiding research programs in cognitive neurobiology and in revealing the nature of cognitive systems. Scientists, after all, not only seek to explain phenomena but also to describe their nature. The use of dynamical systems in explanations of neurocognitive phenomena reflects a deeper commitment to the fundamentally dynamical nature of cognition.

Acknowledgements Many thanks to several anonymous reviewers at multiple different journals. Thanks also goes out to the members of my dissertation committee, who all commented on very early versions of these ideas, including Karen Neander, Felipe De Brigard, Alex Rosenberg, and Walter Sinnott-Armstrong. Special thanks goes to Gualtiero Piccinini and the philosophy of neuroscience reading group at Columbia University.

References

Abrahamsen, A., & Bechtel, W. (2012). From reactive to endogenously active dynamical conceptions of the brain. In K. Plaisance & T. Reydon (Eds.), *Philosophy of behavioral biology* (pp. 329–366). Dordrecht: Springer.

Austin, C. J. (2016). The ontology of organisms: Mechanistic modules or patterned processes? *Biology & Philosophy*, *31*(5), 639–662.

Barack, D. L. (2019). Mental machines. *Biology & Philosophy*, *34*(6), 63.

Bechtel, W. (2002). Decomposing the mind–brain: A long-term pursuit. *Brain and Mind, 3*(2), 229–242.

Bechtel, W. (2008). *Mental mechanisms: Philosophical perspectives on cognitive neuroscience*. Routledge: Taylor & Francis.

Bechtel, W. (2012). Understanding endogenously active mechanisms: A scientific and philosophical challenge. *European Journal for Philosophy of Science, 2*(2), 233–248.

Bechtel, W., & Abrahamsen, A. (2005). Explanation: A mechanist alternative. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences, 36*(2), 421–441.

Bechtel, W., & Abrahamsen, A. (2010). Dynamic mechanistic explanation: Computational modeling of circadian rhythms as an exemplar for cognitive science. *Studies in History and Philosophy of Science Part A, 41*(3), 321–333.

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

Blatt, G. J., Andersen, R. A., & Stoner, G. R. (1990). Visual receptive field organization and cortico-cortical connections of the lateral intraparietal area (area LIP) in the macaque. *Journal of Comparative Neurology, 299*(4), 421–445.

Bokulich, A. (2011). How scientific models can explain. *Synthese, 180*(1), 33–45.

Boone, W., & Piccinini, G. (2016). Mechanistic abstraction. *Philosophy of Science, 83*(5), 686–697.

Britten, K. H., Shadlen, M. N., Newsome, W. T., & Movshon, J. A. (1992). The analysis of visual motion: A comparison of neuronal and psychophysical performance. *Journal of Neuroscience, 12*(12), 4745–4765.

Britten, K. H., Shadlen, M. N., Newsome, W. T., & Movshon, J. A. (1993). Responses of neurons in macaque MT to stochastic motion signals. *Visual Neuroscience, 10*(6), 1157–1169.

Brody, C. D., & Hanks, T. D. (2016). Neural underpinnings of the evidence accumulator. *Current Opinion in Neurobiology, 37,* 149–157.

Campbell, J. (2008). Interventionism, control variables and causation in the qualitative world. *Philosophical Issues, 18*(1), 426–445.

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

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

Chirimuuta, M. (2014). Minimal models and canonical neural computations: The distinctness of computational explanation in neuroscience. *Synthese, 191*(2), 127–153.

Chirimuuta, M. (2017). Explanation in computational neuroscience: Causal and non-causal. *The British Journal for the Philosophy of Science, 69,* 849–880.

Craver, C. F. (2001). Role functions, mechanisms, and hierarchy. *Philosophy of Science, 68*(1), 53–74.

Craver, C. F. (2007a). Constitutive explanatory relevance. *Journal of Philosophical Research, 32,* 3–20.

Craver, C. F. (2007b). *Explaining the brain*. Oxford: Oxford University Press.

Craver, C. F., & Kaplan, D. M. (2018). Are more details better? On the norms of completeness for mechanistic explanations. *The British Journal for the Philosophy of Science, 71,* 287–319.

Darden, L. (2006). *Reasoning in biological discoveries*. Cambridge: Cambridge University Press.

Ding, L., & Gold, J. I. J. C. C. (2011). Neural correlates of perceptual decision making before, during, and after decision commitment in monkey frontal eye field. *Cerebral Cortex, 22*(5), 1052–1067.

Ding, L., & Gold, J. I. J. N. (2012). Separate, causal roles of the caudate in saccadic choice and execution in a perceptual decision task. *Neuron, 75*(5), 865–874.

Ding, L., & Gold, J. I. J. N. (2013). The basal ganglia's contributions to perceptual decision making. *Neuron, 79*(4), 640–649.

Ditterich, J. (2006). Stochastic models of decisions about motion direction: Behavior and physiology. *Neural Networks, 19*(8), 981–1012.

Dupré, J. (2013). I—Living causes. *Aristotelian Society Supplementary, 87*(1), 19–37.

Egan, F. (2017). Function-theoretic explanation. In D. M. Kaplan (Ed.), *Explanation and integration in mind and brain science* (pp. 145–163). Oxford: Oxford University Press.

Felline, L. (2018). Mechanisms meet structural explanation. *Synthese, 195*(1), 99–114.

Franklin-Hall, L. R. (2016). New mechanistic explanation and the need for explanatory constraints. In K. Aizawa & C. Gillett (Eds.), *Scientific composition and metaphysical ground* (pp. 41–74). Berlin: Springer.

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

Glennan, S. S. (1996). Mechanisms and the nature of causation. *Erkenntnis, 44*(1), 49–71.

Glennan, S. (2008). Mechanisms. In S. Glennan & P. Illari (Eds.), *The Routledge companion to philosophy of science* (pp. 404–412). Abingdon: Routledge.

Glennan, S. (2017). *The new mechanical philosophy*. Oxford: Oxford University Press.

Gold, J. I., & Shadlen, M. N. (2007). The neural basis of decision making. *Annual Review of Neuroscience, 30,* 535–574.

Goldman, M. S., Compte, A., & Wang, X.-J. (2010). Neural integrator models. In L. R. Squire (Ed.), *Encyclopedia of neuroscience* (pp. 165–178). Amsterdam: Elsevier.

Hanks, T. D., Kopec, C. D., Brunton, B. W., Duan, C. A., Erlich, J. C., & Brody, C. D. J. N. (2015). Distinct relationships of parietal and prefrontal cortices to evidence accumulation. *Nature, 520*(7546), 220.

Hempel, C. G., & Oppenheim, P. (1948). Studies in the logic of explanation. *Philosophy of Science, 15*(2), 135–175.

Huk, A. C., & Shadlen, M. N. (2005). Neural activity in macaque parietal cortex reflects temporal integration of visual motion signals during perceptual decision making. *Journal of Neuroscience, 25*(45), 10420–10436.

Huneman, P. (2018). Outlines of a theory of structural explanations. *Philosophical Studies, 175*(3), 665–702.

Illari, P. M., & Williamson, J. (2012). What is a mechanism? Thinking about mechanisms across the sciences. *European Journal for Philosophy of Science, 2*(1), 119–135.

Kaplan, D. M. (2011). Explanation and description in computational neuroscience. *Synthese, 183*(3), 339–373.
Kaplan, D. M. (2015). Moving parts: The natural alliance between dynamical and mechanistic modeling approaches. *Biology and Philosophy, 30*(6), 757–786.
Kaplan, D. M. (2017). *Explanation and integration in mind and brain science*. Oxford: Oxford University Press.
Kaplan, D. M., & Craver, C. F. (2011). The explanatory force of dynamical and mathematical models in neuroscience: A mechanistic perspective*. *Philosophy of Science, 78*(4), 601–627.
Katz, L. N., Yates, J. L., Pillow, J. W., & Huk, A. C. (2016). Dissociated functional significance of decision-related activity in the primate dorsal stream. *Nature, 535*(7611), 285–288.
Klein, C. (2017). Brain regions as difference-makers. *Philosophical Psychology, 30*(1–2), 1–20.
Kuhlmann, M. (2014). Explaining financial markets in terms of complex systems. *Philosophy of Science, 81*(5), 1117–1130.
Latimer, K. W., Yates, J. L., Meister, M. L., Huk, A. C., & Pillow, J. W. (2015). Single-trial spike trains in parietal cortex reveal discrete steps during decision-making. *Science, 349*(6244), 184–187.
Levy, A., & Bechtel, W. (2016). *Towards mechanism 2.0: Expanding the scope of mechanistic explanation*.
Lyre, H. (2017). Structures, dynamics and mechanisms in neuroscience: An integrative account. *Synthese, 195,* 5141–5158.
Machamer, P., Darden, L., & Craver, C. F. (2000). Thinking about mechanisms. *Philosophy of Science, 67*(1), 1–25.
Mazurek, M. E., Roitman, J. D., Ditterich, J., & Shadlen, M. N. (2003). A role for neural integrators in perceptual decision making. *Cerebral Cortex, 13*(11), 1257–1269.
Meister, M. L., Hennig, J. A., & Huk, A. C. (2013). Signal multiplexing and single-neuron computations in lateral intraparietal area during decision-making. *Journal of Neuroscience, 33*(6), 2254–2267.
Miłkowski, M. (2016). Explanatory completeness and idealization in large brain simulations: A mechanistic perspective. *Synthese, 193*(5), 1457–1478.
Paz, A. W. (2017). A mechanistic perspective on canonical neural computation. *Philosophical Psychology, 30,* 213–234.
Piccinini, G. (2007). Computing mechanisms. *Philosophy of Science, 74*(4), 501–526.
Piccinini, G. (2010). The mind as neural software? Understanding functionalism, computationalism, and computational functionalism. *Philosophy and Phenomenological Research, 81*(2), 269–311.
Piccinini, G., & Craver, C. (2011). Integrating psychology and neuroscience: Functional analyses as mechanism sketches. *Synthese, 183*(3), 283–311.
Platt, M. L., & Glimcher, P. W. (1999). Neural correlates of decision variables in parietal cortex. *Nature, 400*(6741), 233–238.
Port, R. F., & van Gelder, T. (1995). *Mind as motion: Explorations in the dynamics of cognition*. Cambridge: MIT Press.
Quine, W. V. O. (1960). *Word and object*. Cambridge: MIT Press.
Roitman, J. D., & Shadlen, M. N. (2002). Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. *Journal of Neuroscience, 22*(21), 9475–9489.
Salmon, W. C. (1984). *Scientific explanation and causal structure of the world*. Princeton: Princeton University Press.
Shadlen, M. N., Kiani, R., Newsome, W. T., Gold, J. I., Wolpert, D. M., Zylberberg, A., et al. (2016). Comment on "Single-trial spike trains in parietal cortex reveal discrete steps during decision-making". *Science, 351*(6280), 1406–1406.
Shagrir, O., & Bechtel, W. (2017). Marr's computational level and delineating phenomena. In D. M. Kaplan (Ed.), *Explanation and integration in mind and brain science* (pp. 190–214). Oxford: Oxford University Press.
Shapiro, L. A. (2013). Dynamics and cognition. *Minds and Machines, 23*(3), 353–375.
Shapiro, L. A. (2016). Mechanism or bust? Explanation in psychology. *The British Journal for the Philosophy of Science, 68*(4), 1037–1059.
Silberstein, M., & Chemero, A. (2012). Complexity and extended phenomenological-cognitive systems. *Topics in Cognitive Science, 4*(1), 35–50.
Silberstein, M., & Chemero, A. (2013). Constraints on localization and decomposition as explanatory strategies in the biological sciences. *Philosophy of Science, 80*(5), 958–970.
Strawson, P. F. (1959). *Individuals*. London: Methuen.

Strogatz, S. (2001). *Nonlinear dynamics and chaos: With applications to physics, biology, chemistry, and engineering (studies in nonlinearity)*. Boulder: Westview Press.
Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: The leaky, competing accumulator model. *Psychological Review, 108*(3), 550.
van Gelder, T. (1995). What might cognition be, if not computation? *The Journal of Philosophy, 92*(7), 345–381.
van Gelder, T. (1998). The dynamical hypothesis in cognitive science. *Behavioral Brain Sciences, 21*(5), 615–628.
Wald, A., & Wolfowitz, J. (1948). Optimum character of the sequential probability ratio test. *The Annals of Mathematical Statistics, 19*(3), 326–339.
Walmsley, J. (2008). Explanation in dynamical cognitive science. *Minds and Machines, 18*(3), 331–348.
Wang, X. J. (2002). Probabilistic decision making by slow reverberation in cortical circuits. *Neuron, 36*(5), 955–968.
Weisberg, M. (2013). *Simulation and similarity: Using models to understand the world*. Oxford: Oxford University Press.
Weiskopf, D. A. (2011). The functional unity of special science kinds. *The British Journal for the Philosophy of Science, 62,* 233–258.
Weiskopf, D. A. (2017). The explanatory autonomy of cognitive models. In D. M. Kaplan (Ed.), *Explanation and integration in mind and brain science*. Oxford: Oxford University Press.
Wimsatt, W. C. (1997). Aggregativity: Reductive heuristics for finding emergence. *Philosophy of Science, 64,* S372–S384.
Wong, K.-F., & Huk, A. C. (2008). Temporal dynamics underlying perceptual decision making: Insights from the interplay between an attractor model and parietal neurophysiology. *Frontiers in Neuroscience, 2,* 245.
Wong, K.-F., & Wang, X.-J. (2006). A recurrent network mechanism of time integration in perceptual decisions. *The Journal of Neuroscience, 26*(4), 1314–1328.
Woodward, J. (2003). *Making things happen: A theory of causal explanation*. Oxford: Oxford University Press.
Woodward, J. (2017). Explanation in neurobiology. In D. M. Kaplan (Ed.), *Explanation and integration in mind and brain science* (pp. 70–100). Oxford: Oxford University Press.
Wright, C. D., & Bechtel, W. P. (2007). Mechanisms and psychological explanation. In P. Thagard (Ed.), *Philosophy of psychology and cognitive science*. Amsterdam: Elsevier.
Zednik, C. (2011). The nature of dynamical explanation*. *Philosophy of Science, 78*(2), 238–263.
Zeki, S. M. (1974). Functional organization of a visual area in the posterior bank of the superior temporal sulcus of the rhesus monkey. *The Journal of Physiology, 236*(3), 549.
Zeki, S. (1991). Cerebral akinetopsia (visual motion blindness). *Brain, 114*(2), 811–824.