Phenom Cogn Sci
https://doi.org/10.1007/s11097-017-9536-9

Breaking explanatory boundaries: flexible borders and plastic minds

Michael D. Kirchhoff[1] · **Russell Meyer**[1]

Abstract In this paper, we offer reasons to justify the explanatory credentials of dynamical modeling in the context of the metaplasticity thesis, located within a larger grouping of views known as 4E Cognition. Our focus is on showing that dynamicism is consistent with interventionism, and therefore with a difference-making account at the scale of system topologies that makes sui generis explanatory differences to the overall behavior of a cognitive system. In so doing, we provide a general overview of the interventionist approach. We then argue that recent mechanistic attempts at reducing dynamical modeling to a merely descriptive enterprise fail given that the explanatory standard in dynamical modeling can be shown to rest on interventionism. We conclude that dynamical modeling captures features of nested and developmentally plastic cognitive systems that cannot be explained by appeal to underlying mechanisms alone.

Keywords Metaplasticity · Dynamicism · Mechanism · Intervention · Boundaries · Extended cognition · Material engagement theory

1 Introduction

There is a low likelihood that there is single framework that works optimally for all problems and research avenues in the cognitive sciences. In a different context, this is sometimes referred to as the 'no free lunch theorem' (Wolpert 1996). The reason for is that one explanatory framework or set of basic assumptions might work well in one context but not in another. As a result, the no free lunch theorem implies the need to develop disparate modeling as well as explanatory frameworks when seeking to account for different phenomena, across multiple spatial and temporal scales.

✉ Michael D. Kirchhoff
kirchhof@uow.edu.au

[1] Department of Philosophy, University of Wollongong, Wollongong, Australia

Published online: 18 October 2017

This paper is about explanation in the context of the metaplasticity thesis (Malafouris 2010), developed as part of Malafouris' material engagement theory (2004). The metaplasticity thesis and the material engagement theory are embedded within a larger grouping of views in philosophy of cognition known as '4E Cognition' – embedded, embodied, extended and enactive –.[1] Those who defend such views of cognition argue that cognitive systems should be understood (at least sometimes) as extended. An extended system is a system that is comprised of patterns of cognitive activity distributed over elements of brain, body and environment. The metaplasticity thesis is the view that this extended system is plastic, i.e., with flexible boundaries and with dynamics capable of undergoing transformations over the course of time. It is not only the brain that is plastic on this view. The entire organism is plastic and situated in a plastic network of processes spanning across organism and environment (material, social and cultural) given that networks like these are subject to "continuous re-shaping, re-wiring and re-modelling" across proximal, ontogenetic and phylogenetic timescales (Malafouris 2010, p. 55) Crucially, this implies that biophysical boundaries need not be co-extensive with cognitive boundaries. Moreover, it considers the boundaries that demarcate cognitive systems from non-cognitive elements as "hard-won and fragile developmental and [in the human case] cultural achievements, always open to renegotiation." (Sutton 2010, p. 213). This reveals that the boundaries of cognition are not stable and fixed but inherently fluid and plastic (Kirchhoff 2012). Or, put differently, minds – both their unique properties and distinctive boundaries – are not biologically fixed but develop within a "dynamic bio-cultural system, subject to constant transformations (functional but also structural/anatomical) caused by our ordinary developmental engagement with cultural practices and the material world." (Malafouris 2010, p. 55).

It is always possible to consider coupled systems from different points of view and from different spatial and temporal scales. Ashby, the father of the cybernetics paradigm, captures this precise point in the form of an example: "The chisel in a sculptor's hand can be regarded either as part of the complex biophysical mechanism that is shaping the marble, or it can be regarded as part of the material which the nervous system is attempting to control." (1960, p. 40) We can treat the latter case as an illustration of a clearly defined boundary between the biophysical and the artificial. However, in the former case the boundary is less stable and less clear. Indeed, when viewed functionally rather than anatomically or purely physically, the alleged division between organism and environment becomes much more vague and fragile. Thus we read: "Once this flexibility of division is admitted, almost no bounds can be put to its application … The bones in the sculptor's arm can similarly be regarded either as part of the organism or as part of the 'environment' of the nervous system. Variables within the body may justifiably be regarded as the 'environment' of some other parts [and so on]." (Ashby 1960, p. 40) On this view, what we get is a system within systems (within systems) view of how systems couple to and are embedded within their environments. A nested systems view of

[1] We note only to set it aside that the embedded view does not, strictly speaking, belong to this group of views given that the embedded view has been (in our view, at least) convincingly argued to be mutually exclusive to embodied, extended and enactive views (Rupert 2009). Since this is a minor issue and does not impact on the argument of this paper, we shall not comment on it any further.

this particular kind follows from basic principles in complexity theory (Thiese and Kafatos 2013), developmental systems theory (Griffiths and Stotz 2000), and is the direct focus of material engagement theory (Malafouris 2004).

This should not be taken to imply that there are no boundaries at all between organism and world. Rather, it underpins a central assumption of the metaplasticity thesis; namely, that boundaries of cognitive systems are much more vague, ephemeral and less fixed than assumed (for an analysis of why there no such boundaries with fixed properties, see Kirchhoff 2012).

Extended cognitive systems are intuitively dynamical systems since it is coherent to cast cognitive systems as continuous-time systems (Spivey 2007), complex (Kelso 1995) and as exhibiting emergent dynamics (Silberstein and Chemero 2013; see also Palermos 2014, 2016).[2] The continuous-time aspect of cognitive systems speaks to the idea that cognitive systems are what one might call *voyaging* systems, i.e., they are processes that require for their very existence a dynamic trajectory over time (Kirchhoff 2015). Complex systems are composed of many parts, often with different properties, that interact and interconnect in specific ways. A system is dynamically complex when it exhibits novel and substantially different effects at local and global scales, often with these effects unfolding over different temporal scales – a key example from material engagement theory is the potter working with a clay material. Another example is the formation of Benard cells in physics. In such complex and metastable systems, the global behavior may be difficult to predict even if some of its local scale behavior is not. The idea that cognitive systems exhibit emergent dynamics is entailed by the nature of such systems being self-organising systems given that self-organisation can be shown to induce a partitioning of slow and fast timescales that correspond to global and local dynamics, respectively. The metaplasticity thesis highlights all of these features of cognitive systems.

The idea that cognition emerges from networks comprised of self-assembled couplings between brain, body and environment is implied by the basic principles of dynamical systems theory, given that the mathematical methods of nonlinear dynamical systems theory allow one to consider the properties of random and nonlinear dynamical systems – such as living beings and their environments – and how such properties emerge over time (Chemero and Silberstein 2008; Palermos 2014, 2016). Dynamical systems theory would thus seem to be the perfect explanatory partner for the metaplasticity thesis. In other words, the metaplasticity thesis is a framework for thinking about and exploring the nature of cognition. Dynamical systems theory can provide the metaplasticity thesis with an explanatory dimension.

But the idea that dynamical systems theory (dynamicism, hereafter) is explanatory is contested. Recent years have seen the arrival of a so-called 'mechanistic view of the explanatory virtue of dynamical modeling' (Kaplan and Bechtel 2011; Kaplan and Craver 2011; see also Craver 2007). This is the view that there is no alternative to a mechanistic account with respect to explanation in biology and cognitive science. It states that dynamical models must either be grounded in or wholly reduced to

[2] Note that if the metaplasticity thesis is correct, then the same holds for cognitive systems with a purely neural realisation base. For an excellent treatment of the continuous-time, complex and emergent nature of cognition – both from an extended and a neural perspective – see Malafouris (2010) and Spivey (2007).

mechanistic models by appropriately mapping all the elements of a dynamical model onto the elements of a mechanistic model. Further, defenders of this view state that dynamical models are no alternative to mechanistic models; dynamical models are non-explanatory, descriptive, complements to the mechanistic framework. Silberstein and Chemero (2013) provide the following set of assumptions that some mechanists think follow from their mechanistic framework:

1. Dynamical and mathematical explanations … must be grounded in or reduced to mechanistic explanation (via localization and decomposition) to be explanatory.
2. Dynamical mechanisms are not an alternative to mechanistic explanation but a complement.
3. When dynamical and mathematical models do not describe mechanisms by appropriately mapping elements of the latter onto the former, then they provide no real explanation.
4. At this juncture, dynamical and mathematical models of explanation … not sufficiently grounded in mechanisms have nothing to offer but "predictivism" by way of explanatory force …" (2013, p. 959)

Hence, defenders of the mechanistic view of the explanatory virtue of dynamicism hold that there is no credible alternative conception of explanation on offer than the one entailed by the mechanistic framework.

This is where the no free lunch theorem becomes relevant again. Dynamical systems such as nested and metaplastic systems exhibit global or large-scale features that cannot readily be explained by the activity of only local or micro-scale aspects of a system (Silberstein and Chemero 2013; Kirchhoff 2015; Malafouris 2004; Palermos 2014, 2016; Sporns 2011; Varela et al. 1991). This suggests that dynamical systems theory should have a privileged role – over and above the mechanistic framework – when it comes it addressing and explaining nested and metaplastic systems.

The mechanistic framework starts from considerations about localisation and decomposition. This allows a scientist to break a system into its component parts and understand not only the functions of each part but also how they are related causally, temporally and spatially to one another (Bechtel and Richardson 1993). These heuristics have proven successful for explaining local dynamics of complex systems – a prime case is Craver's (2007) treatment of the action potential. However localisation and decomposition are likely to do less well when our perspective shifts to the global dynamics of complex systems such as generalised synchrony enslaving the dynamics of complex systems. In such systems one can find difference-making elements at the scale of system topologies or connectivity that make a *sui generis* difference to the overall behavior of the system (Sporns 2011). Furthermore, some mechanists emphasise that in order to determine the boundaries of a mechanism it must be possible to locate clearly defined start-up and termination conditions for a mechanism (Machamer et al. 2000). But there are no clearly defined start-up and termination conditions for the boundaries of extended cognitive systems. As Dupré mentions in the context of biological systems, "there is no a priori reason why the process should end, and hence no termination conditions." (2013, p. 29) By induction, it is therefore unlikely that one can capture the characteristics of the metaplastic and voyaging mind by appeal to a mechanistic framework in and of itself. Crucially, a purely mechanistic approach to

cognition cannot easily accommodate the inherently temporal and process-based aspect of extended cognitive systems given that such a mechanistic framework demans a synchronic and proximal perceptive (Kirchhoff 2016).

There are thus reasons to suspect that we need more than just one model to explain the complexity involved in cognitive systems. However, it is our view that the burden of proof does not lie with the mechanist. It lies with those proposing dynamical models – such as the metaplasticity thesis – emphasising non-mechanistic approaches to explanation. The reason for this is that proponents of dynamicism often appeal to a version of predictivism when attempting to ground the claim that dynamical modeling is explanatory. But appeals to nothing but predictive power have been shown to be insufficient for explanation (Kaplan and Bechtel 2011).

The purpose of this paper is therefore to develop an account of how it is possible to understand dynamical models as offering explanations rather than merely descriptions of phenomena. Crucially, in doing so it becomes possible to supplement the overall framework of the metaplasticity thesis with an explanatory program. We shall argue that recent attempts at pulling the mechanistic view and dynamicism apart with respect to explanation is but skin deep. That is to say that the relevant parties to this debate have failed to see what *unifies* their explanatory projects – an *interventionist* account of explanation (Woodward 2003, 2013). We shall argue that this observation has several important implications for the overall debate about the explanatory credentials of dynamicism. Moreover, if this is correct, it opens up the possibility of using dynamicism in order to explain global-scale aspects of nested and developmentally plastic cognitive systems.

The rest of the paper has five sections. Section two sketches the discussion between mechanism and dynamicism with respect to the explanatory credentials of dynamical modeling. Section three is divided into these parts: (a) we start by introducing the interventionist account of explanation; (b) we then show how mechanistic explanation can be understood through the perspective of interventionism; and (c) finally we show that dynamicism can be understood through the lens of interventionism. Section four turns to consider the metaplasticity thesis, especially notions such as organism-niche co-dependency and cognitive boundaries, given an interventionist view of dynamical modeling. Section five sets out the implications that follow from the above. The last section is the conclusion.

2 Mechanisms and dynamicism

There are different kinds of models used in science. But not all of these models are explanatory given the "widely accepted distinction between merely *modeling* a [e.g., system's] behavior and *explaining it*." (Craver 2006, p. 355; italics in original) There are models used primarily as heuristics in experimental design. Some are used to make predictions. Others function to summarize data, and so on. Nevertheless, "some models have an additional property beyond these others: they are explanations." (Craver 2006, p. 355) This raises the question of what a model requires for it is be explanatory.

Dynamical models are usually taken to explain in virtue of enabling a scientist to make predictions. As Stepp et al. say: "dynamical explanations show that particular phenomena could have been predicted, given local conditions and some law-like

general principles … These predictions can be the basis of further experimentations." (2011, p. 432) Unfortunately, grounding explanation in prediction faces a very, very familiar problem. The problem is that nothing about descriptive facts entails anything about explanatory facts. To see this, consider this simple example from Craver (2006). One can reliably use Ptolemy's models of planetary motion to predict the trajectories of the planets in our solar system. But, these models do not explain such trajectories; rather, they merely describe them (Craver 2006, p. 355) Thus, it would appear that *predictivism* is insufficient for explanation. Hence, if dynamicism tethers its explanatory standards to predictivism, it too would be insufficient for explanation.

Not all mechanists deny that dynamical models can have an explanatory role to play in accounting for phenomena in the biological and cognitive sciences. It other words, the choice between *merely* descriptive dynamical models and bona fide explanatory mechanisms is something of a false choice. This suggests that there is room for dynamical models in providing explanations in these scientific disciplines. For example, Kaplan and Craver (2011) argue that dynamical models "in systems and cognitive neuroscience explain (rather than redescribe) a phenomenon *only if* there is a plausible mapping between elements in a model and elements in the mechanism for the phenomenon." (2011, p. 601; italics added) Call this the mechanistic view of the explanatory virtue of dynamical modeling. According to this view, it does not follow, universally, at least, that dynamical models are explanatorily defective, forever doomed to yielding mere descriptions of phenomena. Instead, as Kaplan and Bechtel (2011) point out: "dynamical models do not provide a separate kind of explanation; when they explain phenomena, it is because they describe the dynamic behavior of mechanisms." (2011, p. 440).

Yet, the mechanistic view of the explanatory virtue of dynamical modeling is pointed. It interprets the relationship between mechanistic accounts and dynamicism such that the latter is explanatory only if it is able to represent the causal structures and functions of a mechanism. As the proponents of the mechanistic view of the explanatory virtue of dynamical modeling explicitly state: "What is required to explain a given phenomenon is to identify the responsible mechanism and the conditions under which it is operating." (Kaplan and Bechtel 2011, p. 441) However, this has a substantial implication for dynamicism. On the one hand, if dynamical accounts are explanatory they are so because "they characterize the operations of the underlying mechanism (including how it is related to features of its environment)." (Kaplan and Bechtel 2011, p. 443) On the other hand, when dynamical accounts do not yield a representation of the underlying causal operations of a mechanism, "they fail to provide explanations, whatever their other virtues." (Kaplan and Bechtel 2011, p. 443) On this view, then, dynamicism has explanatory value only if the elements of dynamical models can be mapped onto specific elements of a mechanism.

When assessing the explanatory value of dynamicism we need accept neither predictivism nor the mechanistic view of the explanatory virtue of dynamical modeling. Instead, it is possible to establish that dynamicism has explanatory value given its compatibility with interventionism – just like it is possible to justify the explanatory credentials of mechanisms via interventionism. We now turn to develop the argument in support of these claims.

3 Interventionism: Unifying mechanism and dynamicism

3.1 Interventionism

The central idea of an interventionist account is that any claim about a causal relationship between two variables, X and Y, is a claim about how a particular value or probability distribution of Y would change as a result of an actual or potential intervention on the value of X (or the probability distribution of X).

One suspicion some readers might harbor is that interventionism is best seen as an example of mechanistic explanation (even if not all mechanists explicitly endorse an interventionist approach to explanation). Why think so? Because the idea of intervening on (or manipulating) a particular value or probability distribution of X, say, requires a mechanistic perspective. This is to say that one might think that to intervene on a value of X requires that one is first able to identify the components and their activities in a system for it to accomplish its overall activities. This strategy is known as decomposition (Bechtel and Richardson 1993). Identifying the kind of tasks performed in a system allows for an attribution of function to systemic parts. This strategy is known as localization (Bechtel and Richardson 1993). Given a commitment to decomposition and localization, the idea is that the strategies of decomposition and localization make possible that the system in question can be intervened upon.

But this assumption is not plausible. Indeed, as Campbell (2007) points out, it is a striking feature of interventionism that it makes "*no appeal* to the idea of mechanism. All that we are asking, when we ask whether *X* causes *Y*, is whether *X* is correlated with *Y* under interventions on *X*." Indeed, whether it turns out that there is such a "mechanism linking *X* and *Y* is a further question." (2007, p. 64; both quotes; italics added) It is entirely plausible to endorse an interventionist view of explanation, yet without doing so through the lens of mechanisms. That is, one need not look to parts and components to establish a difference-making account of explanation; rather, it is entirely possible to locate such difference-making effects at the global scale of systemic organization (Woodward 2013).

Here is how Woodward introduces his account: "*X* causes *Y* in some background condition B_i (where background conditions are represented by variables distinct from *X* and *Y*) if and only if there is a possible intervention on *X* in B_i such that if such an intervention were to occur, the value of *Y* or the probability distribution of *Y* would change." (Woodward 2013, pp. 45–46) Following Woodward and Hitchcock (2003), an intervention on *X* is cast in terms of a variable *I* which acts on X. Here the term 'intervention' refers to an idealized, counterfactual intervention on X such that were one to intervene on X it would yield a subsequent effect on the value of Y (or Y's probability distribution).[3] Thus, the interventionist account can be used to answer a range of "*what-if-things-had-been-different* questions." (Woodward and Hitchcock

[3] Woodward provides the following definitions of variables and values, respectively. He says: "variables are properties or magnitudes that, as the name implies, are capable of taking more than one value. Values (being red, having a mass of 1o kilograms) stand to variables (color, mass) in the relationship of determinates to determinables. Values of variables are always possessed by or instantiated in particular individuals or units, as when a particular table has a mass of 10 kg." (2003, p. 39) One could also include to this list the particular values of a particle such as it spin ratio, momentum, polarization, and so on. Or, in systems neuroscience, for example, topological attributes of evolving graphs, time-varying aspects of global brain networks, etc.

2003, p. 4; italics in original) The notion of 'cause' can be understood as a total cause, an actual cause, or a direct cause. For our purposes, it will be enough to introduce the notion of a direct cause. Woodward defines it as follows: "A necessary and sufficient condition for *X* to be a direct cause of *Y* with respect to some variable set *V* is that there be a possible intervention on *X* that will change *Y* (or the probability distribution of *Y*) when all other variables in *V* besides *Y* (or the probability distribution of *Y*) are held fixed at some value by interventions." (2003, p. 55).[4]

3.2 Interventionism and mechanism

In the mechanistic literature, all whom endorse the mechanistic framework does not share in the reliance on intervention. But defenders of the mechanistic view of the explanatory virtue of dynamical modeling do. For example, in *Explaining the Brain*: *Mechanisms and the Mosaic Unity of Neuroscience* (Craver 2007), Craver explicitly endorses Woodward's interventionist model (and so do Craver and Bechtel 2007; Kaplan and Craver 2011). As Craver says: "I use an example from the contemporary neuroscience of learning and memory to defend Woodward's (2003) view that the causal relations in neural mechanisms are relationships that can potentially be used for the purposes of manipulation and control." (2007, p. 63) Indeed, in their review of different notions of 'causings' in the mechanistic literature – conserved quantity accounts, mechanistic accounts, activity-based accounts, and counterfactual accounts – Craver and Tabery (2015) say about the interventionist view that "[u]nlike the views discussed above, this way of thinking about causation provides a ready analysis of explanatory relevance that comports well with the methods for testing causal claims." (Craver and Tabery 2015, p. 7) This is also evident in Kaplan and Craver (2011). They say: "Finally, models that conform to 3 M [model-to-mechanism-mapping] reveals knobs and levers in mechanisms that might be used for the purposes of bringing the mechanism under our control (Woodward 2003)." (2011, p. 613).

There are numerous examples of explanandum phenomena explained mechanistically. One example is the Hodgkin and Huxley (1952) model of the action potential discussed in Craver (2007). Action potentials consist of both rapid and fleeting changes in what is known as the electrical potential difference in a neuron's membrane. This electrical potential difference (measured as the voltage difference across the membrane) is known as the membrane potential. The membrane potential consists of a separation of charged ions on either side of the membrane (Craver 2007, p. 50). As Craver specifies: "In the neuron's resting state, positive ions line up against the extracellular surface. In typical cells, this arrangement establishes a polarized resting potential (Vrest) of –60 mV to – 70 mV […]. In an action potential, the membrane becomes fleetingly permeable to sodium (Na+) and potassium (K+). This allows the ions to diffuse rapidly across the cell membrane." (2007, p. 50) This rapid diffusion changes the mV, in that, the action potential consists of (i) a quick increase in mV to a maximum of +35 mV, which is followed by (ii) a rapid decrease in mV to certain values below the so-called Vrest, followed by (iii) a prolonged or extended after-potential period during which the neuron is less excitable (cf. Craver 2007, p. 50).

[4] There are further conditions to be met such as the requirement that an intervention must be an ideal intervention.We leave such additional conditions aside in the rest of this paper.

Relations of intervention between variables (or their probability distributions) can be expressed by means of directed graphs or equations (Woodward 2003). In a recent treatment, Menzies (2012) provides a systematic analysis of the causal structure of the mechanism underlying the action potential. He does this using both directed graphs and structural equations. Formally, a directed graph consists of an ordered pair (V, E), where V is a set of variables, and E a set of directed edges connecting these variables. A directed edge represents a direct causal relationship between *X* and *Y*. Under simplifying conditions, Menzies represents the directed graph for the action potential mechanism as follows (2012, p. 800) (Fig. 1):

Menzies' (2012) account of the mechanism underlying the action potential is straightforward. In this graph, the components of the mechanisms (i.e., the variables) are axon terminals, vesicles, calcium ion channels, membranes, synaptic clefts, and so on. The values of each variable can be represented as follows (Menzies 2012, p. 800):

A*P* = 1 if action potential arrives at axon; 0 otherwise.
CC = 1 if calcium channels open in cell membrane; 0 otherwise.
P = 1 if calcium sensitive proteins attached to vesicles change shape; 0 otherwise.
F = 1 if vesicles fuse with presynaptic membrane; 0 otherwise.
NR = 1 if neurotransmitters in vesicles are released into synaptic cleft; 0 otherwise.

Figure 2 represents the result of an intervention of CC graphically. This is illustrated by the breaking arrow from AP to CC, which expresses "that the normal causal influence of AP on CC is overridden by the intervention on CC." (Menzies 2012, p. 800).

Two things are important to emphasize. The first is that causal graphs do not strictly speaking provide us with information about the relations between the values of the variables involved. Instead, they represent an intervention on a value of a variable. To get at this kind of information one can add – as Menzies does – structural equations. For example (2012, p. 800):

$$\mathrm{CC} = \mathrm{AP}$$

$$\mathrm{P} = \mathrm{CC}$$

$$\mathrm{F} = \mathrm{P}$$

$$\mathrm{NR} = \mathrm{F}$$

According to Menzies (2012), these "structural equations state that the value of each endogenous variable is determined very simply by the value of one other variable." (2012, p. 800) These equations represent hypothetical or counterfactual causal information and the causal graphs represent the result of an intervention graphically (Menzies 2012, p. 800). They tell us about how the value of P would change were

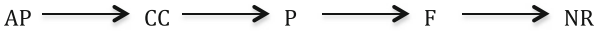

Fig. 1 Causal graph of neurotransmitter example with no intervention (adapted from Menzies (2012, p. 801)

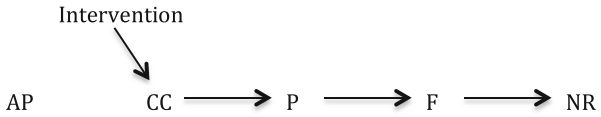

Fig. 2 Causal graph of neurotransmitter example with intervention on CC (adapted from Menzies (2012, p. 801)

one to intervene on the value of CC. The second is that these models emphasize and make plausible the commitment of mechanisms on localization and decomposition. Note that Woodward (2013) Menzies (2012) refer to these notions as modularity. In this context, modularity refers to the idea that one "can associate distinct operations or patterns of behavior or interaction with different subsets of components – each such subset of causal related components continues to be governed by the same set of causal relationships, independently of what may be happening with components outside of that subset, so that the behavior is (in this respect) "intrinsic" to that subset of components." (Woodward 2013, p. 51) As Woodward goes on to say, the "extent to which an explanation or representation is modular is … another dimension which is relevant to whether the explanation is 'mechanical' or exhibits the system of interest as behaving in a machine-like way." (2013, p. 51).

3.3 Interventionism and dynamicism

The claim that those defending the mechanistic view of the explanatory virtue of dynamical modeling are committed to intervention should now be established and uncontroversial.

Once one acknowledges that interventionist explanations fall into the category of 'difference-making' explanations, the assertion that dynamicism exemplifies interventionism should be treated as equally uncontroversial. Woodward has emphasized that difference-making explanations are such that they answer *what-if-things-had-been-different* questions (2013, p. 47). Crucially, for Woodward, this allows that difference-making factors may "occur at many different 'levels' … it may be that the presence of a protein with a very specific structure at a particular concentration is what makes the difference … In other cases, the difference-making factors may be 'higher-level' and less specific – perhaps some effect will be produced as long as a neural network with some overall pattern of connectivity is present …" (2013, pp. 47–48). Once we have the idea that difference-making factors, and subsequent difference-making explanations, can occur at different 'levels' or 'scales,' it is not difficult to show that dynamicism exemplifies interventionism.

The simplest, and best understood, system to use to make this case is convection cells – also called Bénard cells or the Rayleigh-Bénard convection system. Bénard cells form when a thick fluid is heated between two planes or plates in a gravitational field (see Fig. 3):

The emergence of Bénard cells depends on several things such as the type of fluid, its depth, and the temperature gradient. The latter is central. If the temperature gradient is below a certain value (a function of the Rayleigh number), then the fluid will remain stable despite its natural tendency to move given its viscosity and thermal diffusivity. However, when the temperature gradient exceeds a certain critical value (which again is a function of and can be measured by the relevant Rayleigh number), then thermal

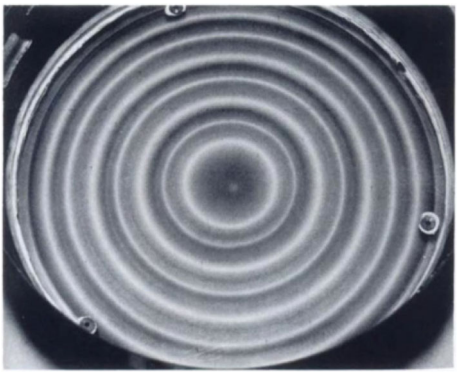

Fig. 3 Convection rolls of alternately larger and smaller section on a circular plate with a radial temperature gradient at the bottom (from Koschmieder 1993, p. 164)

instability occurs. As Chemero and Silberstein put it: as the temperature gradient reaches its critical value, "there is a breakup of the stable conductive state and large scale rotating structures resembling a series of parallel cylinders called Bénard cells are eventually produced." (Chemero and Silberstein 2008, p. 20) All this is to say that one starts to see fluctuations in the density of the fluid given a specific temperature threshold from which large-scale structures – topologies – start to arise.

There are several important things to note about this case, all of which are applicable to dynamical systems – including cognitive and biological ones – in general. First, the Rayleigh-Bénard convection system is an example of a *self-organizing pattern forming system*. This happens when systems spontaneously order themselves without external manipulation of a control parameter (Kelso 1995; Rickles et al. 2007). A second feature of the convection system is that the formation of Bénard cells instantiates as global, relative stable pattern, which 'enslaves' the motion of the individual fluid elements. Chemero and Silberstein make this point as follows: "these large scale structures determine modifications of the configurational degrees of freedom of fluid elements such that some motions possible in the equilibrium state are no longer available." (2008, p. 21) This captures that there is a kind of *circular causality* in the system in which the microscopic dynamics of the fluid elements cause Bénard cells to form while, at the same time, being causally influenced by the macroscopic attributes of Bénard cells. As we noted above, this causal notion of circularity lies at the heart of theorems in the physical sciences such as the slaving principle in physics (Haken 1983). Another aspect of the system is *time-dependence*. That is, the difference between levels of structure in the system is a direct result of the separation of temporal scales given that macroscopic dynamics unfold over longer time scales than does microscopic dynamics. Fourth, in the Rayleigh-Bénard system "the amplitude of the convection rolls plays the role of an *order parameter*." (Kelso 1995, p. 8; italics in original) The order parameter notion refers to a macroscopic (global, systemic) feature of a system. It carries information about how the individual elements of the system are interacting or competing with one another (Rickles et al. 2007, p. 935). Whereas the amplitude of convection rolls functions as an order parameter, the temperature gradient is a *control parameter* (Kelso 1995, p. 7). This concept refers to an external input to the convection system, say, that can be manipulated in order to change the value of the order parameter

and thereby change the macroscopic attributes of the system (Rickles et al. 2007, p. 935). Finally, describing the motion of the fluid under conditions of Rayleigh-Bénard flow requires modeling the motion through a set of *nonlinear* and *coupled differential equations* (Lorenz 1963).

Consider again the ways in which relations of intervention between variables (or their probability distributions) can be represented: by means of directed graphs (as we saw above) or by the use of equations (Woodward 2003). We focus on the latter, given that the use of equations is invoked to model complex systems such as the Rayleigh-Bénard convection system. Here a causal model will consist of an ordered pair (V, E), where V is a set of variables and E is a set of equations. Woodward gives the following illustration: "if $X_1 \ldots X_m$ are all direct causes of Y, then Y may be written as $Y = F_y(X_1 \ldots X_m)$" (2003, p. 42).

A caveat. Woodward's focus is on deterministic causal relations such as when X is a deterministic cause of Y. In dynamical systems theory, a system (or process) is deterministic when and only when it is possible to precisely determine its past and future trajectories from its start up conditions (Rickles et al. 2007). Were the Rayleigh-Bénard convection system to be a deterministic system one could model it by using the above equation, the order parameter Y being a function of $X_1 \ldots X_m$. But the Rayleigh-Bénard convection system is not a deterministic system. Instead it is taken to be a nonlinear and indeterministic system. According to Rickles et al., if a system is indeterministic it "is one without a unique future trajectory, so that the evolution is random [or close to random]." (2007, p. 934) Some might take this to imply that one cannot use Woodward-style interventionism in the system under consideration. Well, luckily things are not as bad as they seem. For Woodward's account can tolerate and thus accommodate indeterministic causation. As he puts it: "The obvious strategy for generating a manipulability theory of indeterministic causation is to replace the references in the deterministic version to manipulations of the value of X that change the value of Y with references to manipulations of X that change the probability distribution of Y." (2003, p. 41).

It is insightful to ask where the order parameter concept comes from. It comes from linear stability analysis in mathematics. In the Rayleigh-Bénard convection system, it should be cast as an "analysis of the equation of motion that describes the [liquid's] behavior … The basic idea is that the initially random starting point can be considered as a [probability distribution] of a whole bunch of different vibratory modes described by this equation." (Kelso 1995, p. 8).

We can now see how dynamical modeling exemplifies an interventionist account of explanation.

Woodward emphasizes that when "causal relationships are … indeterministic, the relevant equations specify how the probability distribution of *Y* [the amplitude of the Bénard cells] will change under manipulation of the right side variables representing direct causes in each equation." (2003, p. 43) One might intervene on the probability distribution of Y by intervening on the control parameter, viz., the temperature gradient, which would result in the system shifting between various phases or regimes. As Rickles et al. observe, this is a "general feature of complex systems; tuning the system's control parameter to a certain critical point results in a phase transition at which the system undergoes an instantaneous radical change in its qualitative features [its velocity, in our case]." (2007, p. 935) Indeed, for complex dynamical systems, informative

interventions are often introduced to global features such as order parameters rather than at the level of microscopic dynamics. Silberstein and Chemero make the same observation but in the contexts of systems neuroscience in which efficacious and informative interventions are performed on topological features such as plasticity, robustness, degeneracy and autonomy (2013, p. 969).

Lorenz equations (1963) encode hypothetical or counterfactual information. So do the simple equations that we explore below in Beer (1995). We know that conceiving of causation in counterfactual terms is at the heart of Woodward's interventionist account. As he puts it: "Causal relationships between variables … carry a … counterfactual commitment: they describe what the response of *Y would be if* a certain sort of change in the value of *X* were to occur." (2003, p. 40) Interestingly, dynamicists such as Stepp et al. are explicit about this. Describing the evolution of a system over time using differential equations yields a description that is "counter-factual supporting." (2011, p. 432).

Finally, in making use of the order parameter concept, it is a signature feature of most dynamical accounts that they involve difference-making features at the network or global scale of systemic operations. It is these difference-making features that are subject to modeling on the basis of differential equation. Support for this claim comes from Woodward himself. He says that "successful dynamical systems and topological explanations have the following properties: … they locate difference-making features for … their explananda, exhibiting dependency relationships between those explananda and factors cited in their explanantia." (2013, p. 63) For these reasons, we conclude that like mechanistic explanation, dynamical modeling is based on and makes use of an interventionist account of explanation.

We now turn to apply some of the above insights on the explanatory credentials of non-mechanistic, dynamical models to the metaplasticity thesis. We start by providing a very broad outline of the thesis by examples. We then turn to consider in some detail how to explain the idea that organisms reflect their environment, and how the environment reflects the organisms that inhabit it through processes of generalised synchrony, which we link back to the idea of global scale difference-making features in dynamical systems.

4 Metaplasticity: Co-dependence, flexible boundaries and intervention

At the core of the metaplasticity thesis is the idea that cognitive systems are nested systems: they are systems within systems (within systems). Hence, a cognitive system can be described as a nearly infinite matryoshka embedding of systems-within-systems (Allen and Friston 2016). Of course, once one adopts such as nested systems-within-systems view, it will always be possible to identify a 'smaller' system relative to which one can say that something else is external to that system – or is not part of that system. Yet despite this multiplicity of systems and systemic boundaries, the key insight is that "we should not simply assume that metabolic boundaries and cognitive boundaries always and everywhere coincide." (Clark 2017, p. 10)[5] Rather than locating the boundaries of cognition at the spinal cord, Malafouris – akin to Ashby before him –

[5] Clark (2017) makes this observation in the context of a different debate. We use it here because it highlights the point that we want to capture; namely, the manifold nature of systems and their boundaries.

stresses that the division between the cognitive system and its niche is anything but clear-cut, with the boundary itself becoming "vague" (Ashby 1960, p. 40).

This already gives us one salient reason for going beyond the mechanistic view of the explanatory virtue of dynamical modeling. Mechanistic explanations, Dupré notes, conceive of the boundaries of a mechanism as stable and clearly identifiable – with specific start and termination conditions (2013). However, if the metaplasticity thesis is on the right track, then cognitive systems are ultimately process-based, and therefore intrinsically temporal. Yet a process has no easily identifiable start and end points (Kirchhoff 2015). A hallmark of dynamicism is its focus on the continuous-time characteristics of cognitive systems. It is therefore a natural alley to any thesis that places emphasis on the flexible and rearrangeable boundaries of cognition such as the metaplasticity thesis and third-wave views of the extended cognition thesis (Kirchhoff 2012; Malafouris 2004; Sutton 2010).

Beer (1995) provides an early dynamical model that dovetails with the interventionist framework. He considers dynamically nested systems, bringing to the forefront the idea that the brain can be modelled as a system within the body, and where a coupled brain-body system can be modelled as a jointly coupled system embedded within a larger system, the environment. Under simplifying conditions, Beer (1995) models an agent and its environment as two continuous-time random dynamical systems, A and E. Formally, the coupling between A and E can be mathematically represented as follows:

$$x_A = A(x_A; S(x_E); u'_A);$$

$$x_E = E(x_E; M(x_A); u'_E)$$

S represents a sensory function from environmental variables to parameters of the agent and M represents a motor function from agent state variables to parameters of the environment. The notation $S(x_E)$ is associated with the sensory inputs of an agent, and $M(x_A)$ with motor outputs. Input arguments u'_A and u'_E correspond to any residual parameters of A and S respectively that are not taken to play a role in the coupling. Feedback plays a crucial role. As Beer emphasises: "Any action that an agent takes affects its environment in some way through M, which in turn affects the agent itself through the feedback it receives from its environment via S [and vice versa]. Thus, each of these two dynamical systems is continuously deforming the flow of the other …" (1995, p. 182) From the kind of coupling highlighted between A and S, Beer goes on to claim that it is possible to view the two coupled dynamical systems assembling a larger system U whose state variables are the *union* "of the state variables of A and S." (1995, p. 183).

Beer's (1995) work on coupled dynamical systems is well known. Crucially, when cast in terms of the metaplasticity thesis, it highlights that given the tight coupling between A and S the dynamics of one system will come to reflect the dynamics of the other and vice versa. A driving assumption of the metaplasticity thesis is that nested dynamical systems in virtue of their inherent plasticity are transformed by one another during time to such an extent that they come to reflect and embody the characteristics of each other. Moreover, it is possible to perturb A, say, by intervening to change the probability distributions of x_A and x_E –

represented in the differential equations above – inducing a qualitative change in the behavior of the output functions (the left side of the equations). This is precisely what one would expect to happen under an interventionist framework focusing on difference-making elements at the scale of order parameters (note that order parameters are probability distributions, i.e., an intervention over distribution of the system parameters).

All this is fairly abstract. However the basic idea is that organisms come to embody regularities of their environment, and vice versa. In other words, organism and environment mutually transform one another as a result of generalised synchrony or co-dependence. That is, by creating their own environments, organisms effectively create themselves in the process. A non-human example of this kind of synergy is the spider and its embedding environment. For example, the spider's morphology, possibilities for action, and so forth, are reflective of the kind of niche it inhabits, while the web and its wider niche reflect the kind of organism that inhabits it.

In the human case, everyday life is ordered by more or less stable socio-cultural patterns, presenting "regularities that arise from everyday practices while at the same time shaping them." (Roepstorff et al. 2010, p. 1051. This kind of patterning not only occurs at the scale of sociocultural practices but also at the more microscopic scale of neurodynamics. That is, regularities at the scale of the sociocultural shape and constrain patterning at the neural scale, viz., that regularities in sociocultural practices come to be reflected in the patterning of cortical connectivity and activity "and in the same way the social practice forms patterns, large-scale brain signals as well as other psychophysical signals generated during particular task performance can be analyzed to expose significant patterning." (Roepstorff et al. 2010, p. 1052).

Consider the following example, based on a mismatch negativity paradigm in cognitive neuroscience. Näätänen et al. (1997) exposed subjects to unattended sounds in certain rhythmic patterns. They found that the primary auditory cortex in the left hemisphere is highly sensitive to changes in predictable sound patterns (1997, p. 432). In the first set of these experiments (1997), Näätänen et al. utilized Finnish and Estonian language speakers because of a relatively small discrepancy between the two languages in terms of vowel structure, except that Estonian vowel space includes an additional vowel, /õ/, not found in Finnish. In the experiment, the speakers were also presented as deviants a prototype of this sound, along with vowels existing in both languages (/o/ and /ö/), and a non-prototypical vowel (located between/e/ and /ö/) (1997, p. 432; Roepstorff et al. 2010, p. 1053). The upshot was that Finnish speakers showed significantly higher mismatch negativity when exposed to prototypical vowels in their native language than when exposed to the Estonian vowel /õ/. This example highlights that the environment embodies regularities and that organisms come to embed such "into their anatomy" (Friston and Stephan 2007, p. 422). This is a process that unfolds over long timescales. If one focuses only on the synchronic timescale of the here-and-now, one looses sight of the mutuality of agent-environment relations. The nice thing about this case is that it highlights the nested view of agents within environments, and how specific regularities within such environments transform how organisms come to perceive their world over time. It is this diachronic mutuality between organism and environment, enabled by porous and ever-evolving cognitive boundaries, that (as we see it) lies at the core of Malafouris' metaplasticity thesis.

Part of the explanation for how this is possible is the emergence of generalised synchrony and therefore covariance in nested and coupled systems (Fig. 4).

Friston and Frith (2015) consider the emergence of perceptual coupling by simulating two birds. They started the simulations with random initial conditions. If the birds cannot hear one another, their respective dynamics will tend to follow independent trajectories, as shown in the left panel in Fig. 4. However, by moving the birds closer to one another – so that they can hear each other – the birds synchronise almost instantaneously, as shown in the right panel of Fig. 4. As Friston and Frith explain: "This is because the listening bird is quickly entrained by the singing bird to correct [anticipate] the … dynamical states generating sensations." (2015, p. 10) This kind of coupling is also known as generalised synchrony. It is ubiquitous in the natural world, yet it highlights that generalised synchrony – driving multiple systems to assemble into a single coherent whole is a macroscopic or global element of dynamically coupled systems. This means that generalised synchrony plays the same role in this context as did the order parameter in the case of convection rolls discussed earlier. Specifically, generalised synchrony carries information about how elements of a system are interacting or competing with one another (Rickles et al. 2007, p. 935). It implies that there is an overall state of stability in the system, and allows us to see how the constituents of a dynamical system can be lesions via intervention into a state of disorder – by effectively minimising their synchrony. This shows that generalised synchrony between two or more systems results in such systems entraining "each other and, effectively, share the same dynamical narrative." (Friston and Frith 2015, p. 12).

Emotion regulation in dyads makes this even more evident. Varga (2015) has recently argued that emotion regulation is not a property of an individual but of a socially extended process. In dyadic emotion regulation between infant and caretaker, the interaction between the two arises because of generalised synchrony. As Varga explains: "Synchrony refers to an unforeseen degree of temporal coordination of non-verbal behaviors of the child and the caretaker. This includes body movements, gaze, vocalizations and affect during the earliest caregiver-child interactions. In such synchronic interactions, there is an

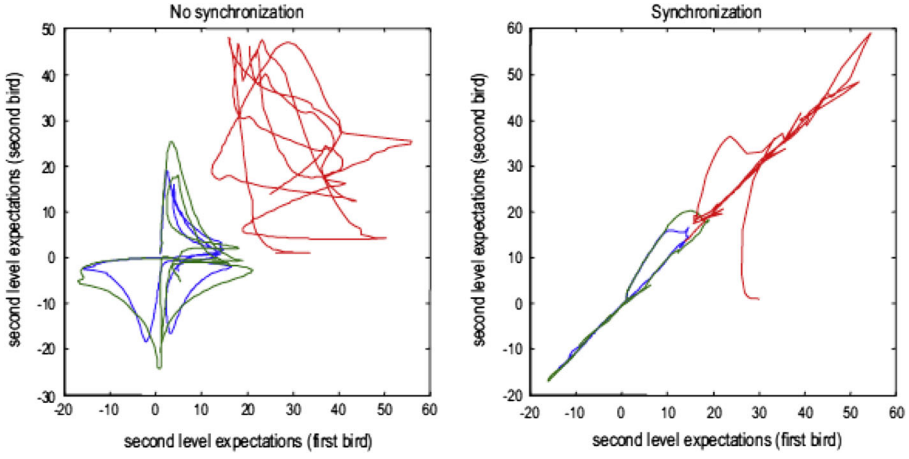

Fig. 4 The left panel shows chaotic and uncoupled dynamics when two birds cannot hear one another, while the right panel shows the generalised synchrony that emerges when the birds exchange sensory signals (adapted from Friston and Frith 2015, p. 11)

emergence and maintenance of non-predetermined synchronic interaction patterns over time, in which caretaker and infant complement each other's states and moderate the level of positive arousal in cooperation." (2015, p. 6).

Having identified generalised synchrony with an order parameter, disorder (or high entropy) can be induced in the caretaker-infant dyad. A classical example is the still face experiment (Tronick et al. 1979). Varga (2015) puts it as follows: "If the continuous, synchronic interaction pattern between the infant and the caretaker breaks down, or if the previously engaged care-taker suddenly puts on a motionless and neutral face, the infant becomes distressed." (2015, p. 7) This reveals several interesting things. The first is that emotion regulation shares the hallmark of dynamical systems; namely, it comes into existence via self-organisation. The second is that dyadic emotion regulation is an ensemble property, not a property of a single individual (Varga 2015), which can be shown to have top-down effects on the individual constituents of the joint system. This follows from the slaving principle at the heart of cybernetics and dynamical systems theory.

The slaving principle puts pressure on any attempt to explain the formation of generalised synchrony from a bottom-up, mechanistic perspective. This is because mechanistic explanation requires an ontology of levels – or minimally an epistemology of levels – partitioning a separation between parts and whole, and vice versa. Yet, consider what Thompson says about the slaving principle: "At this dynamic [scale], the distinction between pre-existing parts and supervening whole has no clear application. One might as well say that the components … emerge from the whole as much as the whole … emerges from the components." (2007, p. 423) If this is correct, then the assembly and re-assembly of nested dynamical systems require more than a mechanistic explanation: it requires a framework that can address macroscale or global features of a system. Dynamical systems theory does so by providing an explanatory account that emphasises difference-making at the global or macroscopic scale in complex systems.

5 Implications and concluding remarks

The overall claim of this paper has several implications for the debate about the explanatory credentials of dynamicism, the mechanistic view of the explanatory virtue of dynamical modeling, and for explanation in the context of the metaplasticity thesis (and more generally for 4E cognitive science).

First, it presents a problem for the mechanistic view of the explanatory virtue of dynamical modeling, which states that dynamical modeling must either be an articulation of mechanistic explanation or a (non-explanatory) guide for such mechanistic explanation. But these claims are not plausible. If the standard for explanation is intervention in both frameworks, then dynamical and mechanistic accounts should be treated in an *even-handed manner*: both have an equal claim to advancing explanations. The difference maker is the form of the explanation given. On the one hand, mechanists, generally speaking, focus on the localization and decomposition of a mechanism, and say that this is what explains a phenomenon. What we get is mechanistic explanation. And for the mechanistic view of the explanatory virtue of dynamical modeling the standard for explanation is interventionism. On the other hand, dynamicists focus

on large-scale, ensemble dynamics represented by equations or models, and say that this has explanatory force. What we get is an explanation of a non-mechanistic kind. And here the standard for explanation is (again) interventionism. Thus, if accounting for a mechanism is explanatory in virtue of exemplifying an interventionist account of explanation, and if accounting for the evolution of a system over time is explanatory in virtue of supporting an interventionist model of explanation, then both frameworks should have an equal claim to explanation. Hence, it is simply not correct to say that dynamical models explain only if they qualify as forms of mechanistic explanation.

Second, certain strands of dynamical modeling focus on prediction, and say that this is what underpins the claim that this form of modeling is explanatory. It is this focus on prediction that leads some to say that dynamical accounts align themselves with covering-law explanation. This view has its share of defenders. But it faces a deep and problematic objection. As Kaplan and Bechtel observe: "Predictivism confronts many well-known shortcomings. For example, by knowing a law-like regularity, one can predict a storm's occurrence from falling mercury in the barometer, but the falling barometer does not explain the occurrence of the storm." (2011, p. 440) We agree. An all too familiar problem with prediction as a rationale for making claims about causal explanation is that it "is overbroad." (Woodward 2003, p. 31) As Woodward explains: "a concern with prediction doesn't explain why we make the distinctions we do between causal and noncausal relationships." (2003, p. 31) But, it need not follow that dynamical accounts are non-explanatory, and therefore merely descriptive. Indeed, once we realize that dynamical accounts need not tether their explanatory standards to predictivism, given that such accounts exhibit interventionism vis-à-vis explanation.

Third, the choice between mere descriptive dynamical models and explanatory mechanisms is a false choice. Instead, we are confronted with a distinction between mechanistic explanation and non-mechanistic explanation. This is consistent with Dupré (2013) and Woodward (2013). Both of these philosophers emphasize that there are key differences between mechanistic (mechanical) and non-mechanistic (non-mechanical) kinds of explanation. Rather than framing the debate as an either/or, the conclusion should be that there is a happy coupling between the two. The mechanists are correct that mechanistic explanation reveals significant and nontrivial aspects of explanation in the sciences. But, there is no need to think that this is the only form of explanation by which to understand phenomena in the biological and psychological sciences. If this is correct, it supports those dynamicists who argue that it is something of a false dilemma to insist that must be accept either mechanistic explanation or dynamical predictivism (Silberstein and Chemero 2013). Drawing on work from the field of systems neuroscience, Silberstein and Chemero (2013) say the following about this work and its implications for explanation: "The brain is modeled as a complex system: networks of … interacting components such as neurons, neural assemblies, and brain regions. In these models, rather than viewing the neurons, cell groups, or brain regions as the basic unit of explanation, it is brain multiscale networks and their large-scale, distributed, and nonlocal connections or interactions that *are the basic unit of explanation*" (2013, p. 963; italics added). Prima facie, at least, this is very different from the mechanistic perspective on explanation.

Finally, the metaplasticity thesis requires a global and distributed unit of explanation. So too does much of the work in 4E approaches to cognition. This suggests a more radical view of the parity view established above between the mechanistic scheme and

dynamicism; namely that given the nested and multiscale dynamics of metastable and metaplastic cognitive systems, non-mechanistic approaches to explanation such as dynamical systems theory is in a position to yield explanations that are (a) genuinely explanatory and (b) without which any attempt at providing merely mechanistic explanations would be impoverished.

Acknowledgements Kirchhoff's work was supported by an Australian Research Council Discovery Project "Minds in Skilled Performance" (DP170102987), a John Templeton Foundation grant "Probabilitizing Consciousness: Implications and New Directions", and by a John Templeton Foundation Academic Cross-Training Fellowship (ID#60708). The opinions expressed in this publication are those of the author and do not necessarily reflect the views of the John Templeton Foundation. Thanks to Lambros Malafouris for inviting us to take part in this special issue and to two anonymous reviewers for insightful comments.

References

Allen, M., & Friston, K. J. (2016). From cognitivism to autopoiesis: Towards a computational framework for the embodied mind. *Synthese, 2016*. https://doi.org/10.1007/s112 29-016-1288-5.

Ashby, R. (1960). *Design for a Brain: The origins of adaptive behavior*. New York: Wiley.

Bechtel, W., & Richardson, R. C. (1993). *Discovering complexity: Decomposition and localization as scientific research strategies*. Princeton: Princeton University Press.

Beer, R. D. (1995). Computational and dynamical languages for autonomous agents. In R. F. Port & T. van Gelder (Eds.), *Mind as motion: Explorations in the dynamics of cognition* (pp. 121–147). Cambridge: MIT Press.

Campbell, J. (2007). An interventionist approach to causation in psychology. In A. Gopnik & L. J. Schulz (Eds.), *Causal learning: Psychology, philosophy and computation* (pp. 58–66). Oxford: Oxford University Press.

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

Clark, A. (2017). How to Knit Your Own Markov Blanket: Resisting the Second Law with Metamorphic Minds. Available online: http://www.x-spect.org/uploads/9/8/1/5/98154170/knittingmarkov8.pdf.

Craver, C. (2006). When mechanistic models explain. *Synthese, 153*, 355–376.

Craver, C. (2007). *Explaining the brain: Mechanisms and the mosaic Unity of neuroscience*. Oxford: Oxford University Press.

Craver, C., & Bechtel, W. (2007). Top-down causation without top-down causes. *Biology and Philosophy, 22*, 547–563.

Craver, C., and Tabery, J. (2015). Mechanisms in Science. *Stanford Encyclopedia of Philosophy*, pp. 1–25.

Dupré, J. (2013). Living causes. *The Aristotelian Society Supplementary Volume, 1*, 19–37.

Friston, K., & Stephan, K. (2007). Free-energy and the brain. *Synthese, 159,* 417–458.

Friston, K., & Frith, C. (2015). A duet for one. *Consciousness and Cognition*, 1–16. https://doi.org/10.1016/j.concog.2014.12.003.

Griffiths, P. E., & Stotz, K. (2000). How the mind grows: A developmental perspective on the biology of cognition. *Synthese, 122*, 29–51.

Haken, H. (1983). *Synergetics: Non-equilibrium phase transition and self-organization in physics, chemistry and biology*. Berlin: Springer.

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

Kaplan, D. M., & Bechtel, W. (2011). Dynamical models: An alternative or complement to mechanistic explanation. *Topics in Cognitive Science, 3*, 438–444.

Kaplan, D. M., & Craver, C. (2011). The explanatory force of dynamical and mathematical models in neuroscience: A mechanistic perspective. *Philosophy of Science, 78*(4), 601–627.

Kelso, S. (1995). *Dynamic patterns*. Cambridge: The MIT Press.

Kirchhoff, M. D. (2012). Extended cognition and fixed properties: Steps to a third-wave version of extended cognition. *Phenomenology and the Cognitive Sciences, 11*, 287–308.

Kirchhoff, M. D. (2015). Extended cognition & the causal-constitutive fallacy: In search for a diachronic and dynamical conception of constitution. *Philosophy and Phenomenological Research, 90*(2), 320–360.
Kirchhoff, M. D. (2016). From mutual manipulation to cognitive extension: Challenges and implications. *Phenomenology and the Cognitive Sciences*, 1–16. https://doi.org/10.1007/s11097-016-9483-x.
Koschmieder, E. L. (1993). *Bénard cells and Taylor vortices*. Cambridge: Cambridge University Press.
Lorenz, E. N. (1963). Deterministic nonperiodic flow. *Journal of the Atmospheric Sciences, 20*, 130–141.
Machamer, P., Darden, L., & Craver, C. (2000). Thinking about mechanisms. *Philosophy of Science, 67*, 1–25.
Malafouris, L. (2004). The cognitive basis of material engagement: Where brain, body and culture conflate. In E. DeMarrais, C. Gosden, & C. Renfrew (Eds.), *Rethinking Mareriality: The engagement of mind with the material world* (pp. 53–62). Cambridge: McDonald Institute Monographs.
Malafouris, L. (2010). Metaplasticity and the human becoming: Principles of neuroarchaeology. *Journal of Anthropological Sciences, 88*, 49–72.
Menzies, P. (2012). The causal structure of mechanisms. *Studies in History and Philosophy of Biological and Biomedical Sciences, 43*, 796–805.
Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Livonen, A., et al. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature, 385*(30), 432–434.
Palermos, O. S. (2014). Loops, constitution, and cognitive extension. *Cognitive Systems Research, 27*, 25–41.
Palermos, O. S. (2016). The dynamics of group cognition. *Minds and Machines, 26*(4), 409–440.
Rickles, D., Hawe, P., & Shiell, A. (2007). A simple guide to chaos and complexity. *Journal of Epidemiology and Community Health, 69*, 933–937.
Roepstorff, A., Niewohner, J., & Beck, S. (2010). Enculturating brains through patterned practices. *Neural Networks, 23*, 1051–1059.
Rupert, R. D. (2009). *Cognitive systems and the extended mind*. New York: Oxford University Press.
Silberstein, M., & Chemero, A. (2013). Constraints on localization and decomposition as explanatory strategies in the biological sciences. *Philosophy of Science, 80*, 958–970.
Spivey, M. (2007). *The continuity of mind*. Oxford and New York: Oxford University Press.
Sporns, O. (2011). *Networks of the brain*. Cambridge: The MIT Press.
Stepp, N., Chemero, A., & Turvey, M. T. (2011). Philosophy for the rest of cognitive science. *Topics in Cognitive Science, 3*, 425–437.
Sutton, J. (2010). Exograms and interdisciplinarity: History, the extended mind, and the civilizing process. In R. Menary (Ed.), *The extended mind* (pp. 189–225). Cambridge: The MIT Press.
Thompson, E. (2007). *Mind in Life*. Cambridge: The MIT Press.
Thiese, N. D., & Kafatos, M. (2013). Complementarity in biological systems: a complexity view. *Complexity, 18*(6), 1–11.
Tronick, E. Z., Als, H., & Adamson, L. (1979). The communicative structure of face-to-face interaction. In M. Bullowa (Ed.), *Before SpHEECh: The beginnings of human communication* (pp. 349–372). Cambridge: Cambridge University Press.
Varela, F., Thompson E., & Rosch, E. (1991). *The embodied mind*. Cambridge: The MIT Press.
Varga, S. (2015). Interaction and extended cognition. *Synthese*, 1–28. https://doi.org/10.1007/s11229-015-0861-7.
Wolpert, D. (1996). The lack of a prior distinctions between learning algorithms. *Neural Computation, 8*, 1341–1390.
Woodward, J. (2003). *Making things happen: A theory of causal explanation*. Oxford: Oxford University Press.
Woodward, J. (2013). Mechanistic explanation: Its scope and limits. *In Aristotelian Society Supplementary Volume*, 87(1), 39–65.
Woodward, J., & Hitchcock, C. (2003). Explanatory generalization, part 1: A counterfactual account. *Nous, 37*, 1–14.