Minds and Machines (2023) 33:153–184
https://doi.org/10.1007/s11023-022-09618-6

Enactivism Meets Mechanism: Tensions & Congruities in Cognitive Science

Jonny Lee[1]

Received: 20 January 2022 / Accepted: 7 December 2022 / Published online: 16 January 2023
© The Author(s) 2023

Abstract
Enactivism advances an understanding of cognition rooted in the dynamic interaction between an embodied agent and their environment, whilst new mechanism suggests that cognition is explained by uncovering the organised components underlying cognitive capacities. On the face of it, the mechanistic model's emphasis on localisable and decomposable mechanisms, often neural in nature, runs contrary to the enactivist ethos. Despite appearances, this paper argues that mechanistic explanations of cognition, being neither narrow nor reductive, and compatible with plausible iterations of ideas like emergence and downward causation, are congruent with enactivism. Attention to enactivist ideas, moreover, may serve as a heuristic for mechanistic investigations of cognition. Nevertheless, I show how enactivism and approaches that prioritise mechanistic modelling may diverge in starting assumptions about the nature of cognitive phenomena, such as where the constitutive boundaries of cognition lie.

Keywords Cognition · Enactivism · Mechanism · Emergence · Downward causation · Dynamical systems

1 Introduction

Enactivism asserts that cognition is to be understood in terms of the dynamic, reciprocal interaction between an organism and its environment (e.g., Varela, Thompson & Rosch, 1991/2017; Thompson, 2007; Stewart et al., 2010; Di Paolo et al., 2017). This indicates several features of effective explanation: the dynamic and emergent nature of cognition is to be recognised whilst reductive explanations that consider only individual agents or their nervous systems are to be rejected. Concurrently, according to new mechanism (or simply 'mechanism'), at least some sciences explain by uncovering the operation and organisation of components that together

✉ Jonny Lee
jonathan.lee@um.es

[1] Department of Philosophy, University of Murcia, Murcia, Spain

constitute, produce or maintain a phenomenon (e.g., Bechtel & Abrahamsen, 2005; Glennan, 2002; Machamer et al., 2000). This mechanistic model has been extended to explanations of cognition (e.g., Bechtel, 2008). Yet with its attention on decomposable and localisable (often neural) components, especially evident in treatments of cognitive neuroscience (e.g., Piccinini, 2020), the mechanistic model may seem to epitomise the reductive, narrow, and non-dynamic approach that enactivists resist. It's little surprise then that enactivists sometimes repudiate a "mechanistic definition of nature" (Gallagher, 2017, p. 23).

This paper investigates the relationship between mechanism and enactivism and attempts to ease tension between the two frameworks, principally by showing that mechanistic explanations are neither necessarily narrow nor reductive and are compatible with assumptions about the dynamic and emergent nature of cognition, per enactivism. I propose, moreover, that whilst enactivists can safely capitalise on mechanistic explanations, enactivism is of heuristic value to those concerned with mechanistic models of cognition. Nonetheless, I demonstrate outstanding differences in starting assumptions between enactivism and approaches that prioritise mechanistic modelling (cf. Lee & Millar, 2022). Whilst existing analyses show how enactivism and mechanism may compete in explaining particular phenomenon (e.g., Herschbach, 2012, on social cognition), or how branches of enactivism are actually best construed as offering mechanism sketches (e.g., Vernazzani, 2014, 2019, on sensorimotor theory)—both of which I will draw on—this paper adopts a broader approach to comparing the frameworks by targeting their elementary assumptions about the nature of explanation.

Textual evidence for perceived tension between enactivism and 'mechanistic' or 'mechanical' theories is found in at least five overlapping guises: (1) general contrasts between enactivism and mechanistic approaches (e.g., Thompson, 2007); (2) stated differences between organisms and 'machines' or 'mechanical' systems (e.g., De Jesus, 2016); (3) purported independence of mechanistic and dynamical explanations (e.g., Chemero & Silberstein, 2008), the latter of which is appropriated by enactivists (e.g., McGann et al, 2013); (4) apparent strain between mechanism and holistic, non-reductive features of causal dynamics stressed by enactivists, such as emergence and top-down causation (e.g., Varela, Thompson & Rosch, 1991/2017); and (5) proposed incompatibility between (some versions of) mechanism and the Merleau Ponty inspired phenomenology frequently assimilated by enactivism (e.g., Sheredos, 2021).[1]

The relationship between enactivism and mechanistic explanation thus concerns a diverse terrain of ideas that cannot be traversed in one undertaking. To begin

[1] Not all enactivists explicitly challenge the value of mechanistic explanation, and appeals to mechanisms can be found in enactivist texts, for instance, when Varela et al. (1991/2017) write of "'the body as a lived, experiential structure and the body as the context or milieu of *cognitive mechanisms*" (1991/2017, p. lxii, emphasis added). However, the incorporation of mechanisms within enactivist explanations does not diminish the need for investigation but invites it, given inconsistent attitudes across the literature. Moreover, 'mechanism' is somewhat ambigious, and resistance to so-called mechanism may not always target the narrow sense of the term intended by new mechanists. Again, I take this to only further motivate an analysis that can clarify the relationship between enactivism and mechanism in the technical sense developed by new mechanists.

making inroads, I focus on three areas related to the narrowness/breadth and reductive/holistic dimensions of explanation in cognitive science: (1) the appropriate unit of analysis for explaining cognitive phenomena; (2) the possibility of emergence and downward causation; and (3) the role of dynamical descriptions. This taxonomy is not exhaustive but represents a significant territory where enactivism and mechanism may seem to part ways. Whilst all three topics have been discussed elsewhere, to my knowledge, they have not been explicitly leveraged to offer a global comparison of enactivist and mechanistic explanations. An ecumenical project such as this runs the risk of antagonising parties on both sides, but I hope the path laid through the conceptual landscape will help orient explorers of these topics, regardless of their commitments.

The paper proceeds as follows. Section 2 outlines enactivism and mechanism, highlighting key assumptions made about explanation. Section 3 provides a classification of issues where conflict may arise between these frameworks. Section 4 builds on the preceding discussion to explore what benefits enactivism and mechanism afford each other, focusing on the heuristic value of enactivism for mechanism, and closes by gesturing toward some still unresolved tension.

2 Sketching Enactivism & Mechanism

This section introduces enactivism and mechanism, emphasising principles pertaining to explanations of cognition. Both accounts are multi-faceted traditions with internal schisms among devotees. Enactivism, in particular, is not a single unified approach but a set of traditions that share a history and thematic core (Ward et al., 2017). This paper will concentrate on autopoietic enactivism as presented by, for instance, Varela et al. (1991/2017), Di Paolo (2005), and Thompson (2007).[2] I take autopoietic enactivism to be the most expansive type of enactivism on the market, the most common affiliation of self-identifying enactivists today, and most pertinently, the variety of enactivism where animosity with mechanism is most apparent. One might make a further distinction between classical 'autopoietic theory' and contemporary enactivism, equivalent to 'Maturanian enactivism' (after Humberto Maturana) and 'Varelian enactivism' (after Francisco Varela), where the latter shares its roots with but is non-identical to the former (Villalobos, 2013; Villalobos & Silverman, 2018; Villalobos & Ward, 2015). Where it matters for expository purposes, I will defer to the so-called Varelian variety of autopoietic enactivism, also known as 'canonical enactivism' (e.g., Villalobos & Ward, 2016).

2.1 Enactivism

Enactivism depicts cognition as emerging through the dynamic interaction between an active organism and its environment. At the centre of this picture is an

[2] Major enactivist offshoots include 'radical enactivism' (Hutto & Myin, 2012) and 'sensorimotor theory'. (O'Regan & Noë, 2001). For a history and outline of these branches, see Ward et al. (2017).

'autonomous' system coupled with its environment. At first pass, autonomy refers to a type of system that continuously produces and maintains the conditions for its ongoing existence, whilst two systems are coupled when the state of one forms a parameter of the other, and vice versa (e.g., Thompson, 2007, p. 45). The process of 'autopoiesis' exemplifies autonomy, being the capacity of an organism to maintain and reproduce itself by generating and maintaining its own parts and processes. Specifically, for enactivists, an autonomous system is one constituted by a network of processes that recursively depend on one another to produce the very processes themselves, and in doing so, realise the system as a unified individual (e.g., Thompson, 2007, p. 37).

Autonomy is the foundation for the enactive conception of cognition: "what makes living organisms cognitive beings is that they embody or realize a certain kind of autonomy—they are internally self-constructive in such a way as to regulate actively their interactions with their environments" (Thompson & Stapleton, 2009, p. 24). Explicating autonomy, enactivists often appeal to the notion of 'operational closure' which refers to a closed network of enabling relations i.e., a network of processes that simultaneously depend on each other for their existence (for discussion of operational closure within the development of enactivism, see Barandiaran, 2017). For example, in forming a (rather complex) closed network, a living cell's processes mutually enable each other, as when a membrane enables metabolic processes through appropriate spatial containment which in turn sustains the membrane through repair processes (Di Paolo & Thompson, 2014). In turn, enactivists sometimes stress the importance of precariousness in the operationally closed networks that realise autonomy: were the enabling relations to cease, the processes which constitute the network would also cease (e.g., Di Paolo & Thompson, 2014; Bich & Arnellos, 2012).

The capacity of organisms to recursively maintain their biological organization, or 'autopoiesis', forms the exemplar case of autonomy and is the most basic form of self-constitution (e.g., Ruiz-Mirazo & Moreno, 2004). In turn, all organisms (*qua* autopoietic systems) count as cognitive—a theme stretching back to enactivism's roots in the 'Santiago theory of cognition' (e.g., Maturana, 1970; but see Barandiaran, 2017, for complications). As Maturana and Varela (1980) declared: "Living systems are cognitive systems, and living as a process is a process of cognition. This statement is valid for all organisms, with or without a nervous system" (p. 13). Enactivists also identify other forms or levels of (interconnected) autonomy, such as neurological, immunological, and sensorimotor. Our discussion will operate with enough generality that the relation between these forms of autonomy need not overly concern us.

In addition to autonomy, many enactivists stress the importance of 'adaptivity', the capacity of an autonomous system to regulate their (existence-sustaining) operationally closed processes by adjusting their interaction with the environment, thus further promoting the viability of the system (e.g., Di Paolo & Thompson, 2014). Supplementing autonomy with adaptivity provides us with the concept of 'adaptive autonomy' (e.g., Thompson & Stapleton, 2009), whereby a system regulates its interactions with the environment and hence its viability (the conditions under which it can persist as an individuated and unified system).

Adaptive autonomy implies an interconnectedness between agent-environment interaction and that agent's persistence. It also provides the conditions for 'sense making', in which adaptive autonomous systems (agents) evaluate aspects of their environment as beneficial or harmful, or as promoting or endangering viability. As Di Paolo & Thompson write, "An adaptive autonomous system produces and sustains its own identity in precarious conditions, registered as better or worse, and thereby establishes a perspective from which interactions with the world acquire a normative status" (2014, p. 73). Sense-making, rooted in adaptive interactions that advance or hinder the persistence of an autonomous system, gives the environment 'meaning' for the system.

Establishing itself as an alternative to cognitivist and intellectualist approaches (that stress computation, representation, and inference), enactivists maintain that cognitive systems are not faced with reconstructing task-neutral information from a 'pre-given' world but with managing their interactions relative to their conditions of viability: ''cognition in its most encompassing sense consists in the enactment or a bringing forth of a world by a viable history of structural coupling'' (Varela et al. 1991/2017, p. 205). By rejecting a pre-given world that needs to be reconstructed through internal, representational processes, enactivism is taken to undermine a cognitivist approach that explains cognition by appealing to internal 'mental' or 'cognitive' representations—at least in a traditional guise. Assuming computation to require representation, enactivists also typically reject the idea that cognition is computational (e.g., Varela, Thompson & Rosch, 1991/2017; but see Villalobos & Dewhurst, 2017). This is consistent with the widely shared assumption across a spectrum of views that computation is semantic (e.g., Sprevak, 2010; but see Piccinini, 2008), and representational theories derive their explanatory power through the posited computational operations that process representational states (we return to this in Sect. 4).

Subverting the primacy of computation and representation, enactivism aims "to determine the common principles or lawful linkages between sensory and motor systems that explain how action can be perceptually guided in a perceiver-dependent world" (Varela, Thompson, and Rosch, 1991/2017, p. 173). Such lawful linkages can be unpacked in terms of sensorimotor coupling, where bodily variables are coupled with environmental variables. Here, bodies and environments are coupled when some parameters governing equations in one of the systems depend on the state of the variables of the other system (e.g., McGann et al., 2013, p. 204). The purported explanatory power of sensorimotor coupling is emphasised by 'sensorimotor theory' (e.g., O'Regan & Noë, 2001). Sensorimotor theory explains perception through a broadly enactive lens, stressing that perception consists of an agent's interaction with the world, through implicit attunement to 'sensorimotor contingencies', that is, the systematic ways our bodily actions induce sensory changes (I take enactivism, as defined above, to incorporate sensorimotor theory but sensorimotor theory not to entail enactivism). Such contingencies include, for example, the ways that retinal stimulation changes in proportion to eye rotation.

The notion of lawful linkages may seem to indicate explanations of a form captured by the 'covering law' model (Hempel, 1965). Briefly, the covering law model says that explanations operate by showing how an explanandum (what is being

explained) is derivable via a logical argument (either deductive or inductive, in its most plausible guise), which contains at least one 'law of nature' (roughly, some fundamental regularity). This is typically taken to be a rival of the mechanistic model of explanation, and causal accounts more generally (for an introduction, e.g., see Okasha, 2002). The prioritising of laws also aligns with the appeal of dynamical systems theory (DST) which provides enactivists with the mathematical tools to describe sensorimotor couplings and is sometimes interpreted as offering covering law explanations (cf. Bechtel, 1998). We will return to DST and its form of explanation in Sect. 3.3.

2.2 Mechanism

Mechanism proposes a model of explanation in multiple disciplines including cognitive science, forming an alternative to the covering law model of explanation. An evolution of earlier causal-mechanical models of explanation, which state that a phenomenon is explained by its cause, mechanistic explanations explain *why* by showing *how* (e.g., Bechtel & Abrahamsen, 2005). In this way, a cognitive capacity is explained by understanding its underlying mechanism.

A mechanism is a composite of physical entities (components), that are organised (spatially and temporally), such that their operations or activities (types of causes) produce, constitute or maintain a phenomenon. As Bechtel and Abrahamsen (2005) summarise: "A mechanism is a structure performing a function in virtue of its components parts, component operations, and their organization. The orchestrated functioning of the mechanism is responsible for one or more phenomena" (p. 423). For example, the heart is a mechanism for blood circulation because the properties and organisation of its components (valves, atrium, aorta etc.) collectively produce the pumping of blood around the body.

As the heart exemplifies, mechanisms are typically decomposable because we can identify the organised components (parts and their mutual relations) and operations performed by those components that comprise the mechanism (though there may be exceptions; see Povich & Craver, 2018). The organisation of components and their activities are central to understanding how a mechanism realises a phenomenon. Components are arranged by their spatial, temporal and organisational properties. Investigating the location, size and orientation of components (spatial properties), as well as the order, rates and duration of their activities (temporal properties), in conjunction with any general organisational relations such as positive or negative feedback (organisational properties) is, therefore, key to mechanistic explanation. As this emphasis on organisation reveals, in discovering *how* something works by understanding the components that constitute it, a system's parts are not viewed in isolation but in interaction.

Component parts and processes and the mechanisms they constitute are individuated relative to the phenomenon they realise: what counts as a component is determined by what entities/activities collectively function to realise a phenomenon. As mechanists are fond of saying, there are no mechanisms *as such*, only mechanisms *for* phenomena (Glennan, 2002). Thus, identifying and characterising phenomena

is of paramount importance to mechanistic explanation. Nevertheless, how scientists understand a phenomenon can change throughout an investigation; phenomena are frequently not fixed ideas but 'moving targets' (Kronfeldner, 2015). As Bechtel & Richardson (1993) note, investigators will sometimes 'reconstitute' (or recharacterise) the phenomenon being studied over the course of attempting to identify a mechanism. Thus, one's understanding of a phenomenon and one's understanding of a mechanism evolve in parallel. For example, as Bechtel (2008, chapter 2) explores in some detail, memory was once conceived as a more-or-less veridical recapitulation of past events (at least when functioning properly). However, over decades of research, memory has come to be characterised as a constructive activity that is highly interwoven with other mental activities which it was previously demarcated from (various types of memory are also now distinguished e.g., episodic versus semantic, and short term versus long term). What is interesting, for our purposes, is how investigating mechanisms can contribute to our very conception of mental or cognitive phenomena.

Roughly speaking, some entity or process x partially constitutes a phenomenon P if x is an internal part of P. To be an internal part of a mechanism, something must be spatiotemporally contained *within* the mechanism. As such, to explain something constitutively is to explain the capacity or behaviour of a system in terms of its internal causal structure—the parts, processes, and relations that comprise the system "from the inside" (e.g., Piccinini, 2020).[3] To anticipate a later point, we can think of a mechanism as neither separate from its organised, interacting parts nor identical to them. Instead, it is an invariant that may persist despite some transformations in its constitutive elements (my watch is not independent of its organised parts but also persists despite myriad micro and macro physical changes in those parts), with its 'higher-level properties' being *aspects* of their realizers ('lower-level properties'). Notice too that constitutive parts do not exhaust what is necessary for a phenomenon. For example, exposure to sunlight is a cause of seed growth but such exposure does not (mechanistically) constitute the growing seed. From this, we can extract a general principle: mechanistic explanation concerns how organised parts realise a phenomenon under particular circumstances.

Mechanisms are sometimes classified as constituting, producing or maintaining, relative to the phenomenon they explain.[4] Constitutive mechanistic explanation understands a phenomenon in terms of its internal, lower-level realiser, as we have seen. Productive mechanistic explanations, meanwhile, explain some product as the result of a causal sequence (e.g., protein synthesis). In fact, 'production' and 'constitution' needn't reflect exclusive kinds of relations but different aspects of the same mechanism/phenomenon; for instance, we may look at the underlying parts of a single step in a production mechanism or the sequence of steps leading to a product

[3] This does not preclude highly distributed mechanisms, where components are spatiotemporally spread, say, across organism and their environment (see Sect. 3.1). Thus, what counts as 'internal' to a given mechanism might not be obvious or correspond to our folk ontology. I thank Joe Dewhurst for pressing me on this point.

[4] I take the difference between constitutive and productive mechanisms to match the common distinction courtesy of mechanists between 'constitutive' and 'etiological' explanations (e.g., Kaiser & Krickel, 2017).

within a constitutive part. Whether constitution or production is more relevant will depend on which aspect of a mechanism-phenomenon matters most given explanatory context.

We noted that mechanisms constitute, produce *or* maintain a phenomenon, but it may be more accurate to say that maintenance mechanisms are a special case of either productive or constitutive mechanisms (the latter involving the continuous behaviour of a whole mechanism), depending on whether the phenomenon being explained is a stable point or ongoing process (for discussion, see Kästner, 2021). In any case, the cyclic organisation of many mechanisms is crucial for understanding how mechanisms explain continuous biological phenomena involving (often many) feedback loops (e.g., circadian rhythms) and is likely key to any mechanistic explanation of cognition that will satisfy enactivists, given the role of ongoing self-maintenance and organism-environment coupling for adaptive-autonomy. The important lesson is that mechanisms need not be linear or temporally closed (i.e., defined by an endpoint).

Mechanists have proposed 'mutual manipulability' as a generic test of constitution, as introduced above (Craver, 2007).[5] This will be relevant to our analysis below. The gist is that intervening on a component affects the behaviour of the mechanism as a whole, whilst intervening on the behaviour of the mechanism as a whole affects the component (as the mechanism partially consists of the component). In Craver's original formulation:

> (i) x is part of S; (ii) in the conditions relevant to the request for explanation there is some change to X*'s* φ-ing that changes S's ψ-ing; and (iii) in the conditions relevant to the request for explanation there is some change to S's ψ-ing that changes X's φ-ing. (Craver, 2007, p. 153)

Take Craver's example of a word stem completion task, in which subjects must complete word stems of previously given words. Heart rate, it transpires, affects performance on such a task. This might suggest that heart rate is a constitutive component of the mechanism underpinning the word stem completion capacity. However, in most circumstances, undertaking the world completion task itself does not affect heart rate. The invariable change is asymmetric. Therefore, whilst heart rate may make a difference, it is not constitutive of the mechanism underpinning the word stem completion capacity. Of course, we should allow that, in practice, finding ways to isolate changes in both X's φ-ing and S's ψ-ing may be difficult in a complex, dynamical system, however, mutual manipulability at least provides us with the grounds for a somewhat well-defined measure of constitution, in principle. We will return to this in Sect. 3.1.[6]

[5] Mutual manipulability has been hotly debated in the mechanism literature (for some sample concerns, see Leuridan, 2012). This had led to some revision of the original presentation (notably, Krickel, 2018). Most of the nuances emerging from these developments will not impinge on our discussion.

[6] Povich and Craver (2018) suggest that mutual manipulability is sufficient but perhaps unnecessary for composition because there may be parts that do not change at points in its operation or are redundant and so may be affected without any corresponding change in the mechanism as a whole. At minimum, mutual manipulability may still provide a well-defined sufficiency test for (mechanistic) componency.

3 Comparing Explanatory Principles

This section presents three topics that can be used to compare enactivist and mechanist attitudes towards explanations of cognition: the appropriate unit of analysis; the possibility of emergence and downward causation; and the role of dynamical descriptions. These issues crosscut and overlap, hence separating them belies the full complexity of the situation. Nevertheless, isolating these topics affords an otherwise intractable comparison. This taxonomy is not intended to be exhaustive, nor reflect the assumptions of every self-identifying enactivist and mechanist. Rather, it is intended to capture three of the most prominent aspects of explanation on which vanilla varieties of enactivism and mechanism can be compared.

The apprehension haunting enactivism and mechanism can be summarised simplistically but usefully with an apparent dichotomy: wide and holistic (enactivism) vs narrow and reductive (mechanism). Notice the different dimensions here; wide/narrow and holistic/reductive. 'Wide/narrow' refers, roughly, to cognitive boundaries whereas 'holistic/reductive' refers, roughly, to relations between sets of entities at different 'levels' (a term which itself requires unpacking). These should not be conflated: if the mechanistic model implies reductive explanations, it does not follow that it implies narrow explanations. The topics explored below can be mapped to these dimensions: 'unit of analysis' concerns cognitive boundaries; 'emergence and downward causation' concerns relations between levels; and 'dynamical descriptions and reciprocal causation' straddles both. We also ought to avoid confusing the deficiency of an explanation with its being erroneous. Any limits in the mechanistic model of explanation, for instance, could signify (1) an impoverished understanding of mechanistic explanation in the current literature—pointing to the need for further examination—and/or (2) the need to complement but not exclude mechanistic explanation with some other form of non-mechanistic explanation. For example, if consciousness cannot be explained mechanistically, as some enactivists suggest (e.g., Fuchs, 2017), it does not follow that no cognitive phenomena can be explained mechanistically. The same applies (*mutatis mutandis*) for enactivist explanations; mechanists should not confuse enactivism's limits with its falsity. Care is therefore required to avoid any accidental disposal of babies with bathwater.

3.1 The Unit of Analysis

The question of what scale, level or collection of entities must be studied to understand cognition will run throughout the remainder of the paper. However, it will be fruitful to begin our comparison of enactivist and mechanist attitudes to explanation with some preliminary discussion of the appropriate unit of analysis in cognitive science. Different accounts of cognition assume different units of analysis, meaning, the parts and processes of the world that must be considered to explain cognitive phenomena. Not all explanations must target the appropriate unit in its totality. For example, if cognition must be understood in terms of the bodily activities of organisms coupled with their environments, it does not follow that neuroscience cannot study the signalling properties of neurons, only that understanding cognition fully

requires situating those properties within the embodied and embedded agent. As we shall see, this topic overlaps with concerns regarding cognitive boundaries i.e., whether cognition is constitutively restricted to the organism or encompasses parts of the organism's environment. If some social cognitive competence is partially constituted by other agents, for instance, then those agents must be considered within a complete explanation of that phenomenon.

Enactivists typically reject explanations of cognition that take the ultimate unit of analysis as the individual agent or their brain. Instead, they assume the organism plus environment—or more accurately, the organism-environment totality—must be considered. For example, Gallagher (2017) writes, "From a third-person perspective the organism-environment is taken as the explanatory unit" (p. 6). Others have drawn parallel conclusions in rejecting 'methodological individualism' which takes the individual agent as the appropriate target of study (e.g., Froese & Di Paolo, 2011). This is a corollary of the more commonly discussed notion (in the philosophy of mind) of 'methodological solipsism' (Fodor, 1980), the view, roughly, that cognition should be studied in abstraction from the agent's environment.[7] This is associated with the classical 'representational theory of mind' in which representations, in the form of computational states, are of primary explanatory importance, and individuated only by internal relations.[8]

Enactivists not only question the necessity of computation/representation but propose that organisms engage in reciprocal relations with their environment such that agent-environment coupling is crucial to explanations of cognition (e.g., Barandiaran, 2017). These relations invite dynamical descriptions, examined further in Sect. 3.3. For now, notice the contrast between explanations in terms of internal computational mechanisms and agent-environment reciprocity. We might, for instance, compare explanations of social cognition in terms of whole persons and their dynamic interactions with each other (an enactivist explanation) and explanations in terms of brain-bound simulation mechanisms (a mechanistic explanation) (Herschbach, 2012).

For many enactivists, interactions between multiple agents literally constitute social cognition so cannot be fully explained by appealing to mechanisms contained within the nervous system. In other words, the phenomenon of social cognition encompasses multiple agents. Offering such a constitutive view of multiple agents in social cognition, De Jaegher et al., (2010) write that "social cognition is not reducible to the workings of individual cognitive mechanisms", and "interactive

[7] The term 'methodological individualism' is more common in social science, referring to the idea that social phenomena are explained by the mental states and motivations of individual agents. Within social science, there is a parallel criticism to one found in cognitive science, namely, that 'individualistic' explanations over privilege lower-level analyses (e.g., Currie, 2001). Curiously, a shift from covering law to mechanistic models of explanation occurred in social science around the same time as cognitive science, and may help resolve issues pertaining to the unit of analysis (e.g., Hedström & Ylikoski, 2010).

[8] The argument for methodological solipsism, as presented by Fodor (1987), does not begin by precluding the environment playing an individuating role. Rather, it assumes individuation occurs on the basis of causal powers and conjoins this with the claim that causal powers supervene on local 'microstructure', which in the case of cognition, happens to be neural structure (e.g., Fodor, 1987, p. 44). For extended discussion, see McClamrock (1991).

processes are more than a context for social cognition: they can complement and even replace individual mechanisms" (p. 441). Explaining social cognition might require appeals to synchronization, for example, captured by the mathematical tools of dynamical systems theory; a dialogue between two people, say, involves synchronization between speech and bodily movements (De Jaegher & Di Paolo, 2007; cf. Herschbach, 2012).

We must be careful to differentiate two views when assessing claims that social cognition is not reducible to 'individual cognitive mechanisms': (1) a view according to which social cognition is explained in terms of interactions that encompass mechanism components outside an agent's body, or at the very least dynamical interactions between mechanisms inside agents and (non-constitutive) parts of the world, versus (2) a view according to which social cognition is explained without *any* appeal to mechanisms. Only view (2) is at odds with the mechanistic model. This brings us to a central claim of this section: nothing within mechanism alone implies cognition takes place wholly within the organism. Thus, mechanism does not preclude the 'organism-environment' as the appropriate unit of analysis. However, as we shall see, the contingent nature of the environment's constitutive role may persist as an outstanding divergence between enactivism and an approach that prioritises mechanistic modelling.

Mechanistic explanation in fact provides a constructive lens through which to view the possibility of the environment partially constituting cognition given its intrinsic indifference to the spatial location of components.[9] Mechanism allows *but does not predetermine* that the environment constitutes cognitive phenomena, qua mechanistic constitution (we will return to this contrast below). Many mechanists have embraced the possibility of cognitive mechanisms extending into the environment (e.g., Kaplan, 2012; Miłkowski et al., 2018; Zednik, 2011). Especially influential has been Kaplan's (2012) appropriation of the generic mutual manipulability criterion, introduced above, to substantiate claims about extended cognition. Recall that, in Craver's words, "a component is relevant to the behavior of a mechanism as a whole when one can wiggle the behavior of the whole by wiggling the behavior of the component *and* one can wiggle the behavior of the component by wiggling the behavior as a whole" (2007, p. 153; original emphasis). Kaplan's insight is that this empirically tractable criterion can be leveraged to transform questions about whether the environment is part of cognition into questions about whether it's part of a cognitive mechanism, discoverable via the relationship of mutual manipulability. As Kaplan notes, neuroscientists "gain confidence that they have discovered the real components in a neural mechanism" when two experimental strategies are combined: "engaging subjects in task performance while monitoring changes in putative

[9] Really, there are two different kinds of 'environment': the organism's environment (things outside the organism) and the mechanism's environment (things outside the mechanism). These need not coincide because mechanisms may be a subset of the whole organism, or themselves contain the whole organism as a subset. Indeed, mechanistic language may assist enactivists in making sense of why organisms are autonomous entities strictly demarcated from their environment and yet cognition, in a sense, 'extends' into the environment; organisms do not leak into their environment, but the mechanisms for cognition are not restricted to organism boundaries (for possible pushback, see Villalobos & Palacios, 2021).

component(s), and manipulating putative component(s) while detecting changes in overall task behavior" (p. 558). Manipulating gravitational forces affects performance on a reaching task (following Fisk, 1993), but interventions on the reaching task do not affect gravitational forces; by contrast, manipulating eye saccades affects performance on certain memory-intensive tasks whilst interventions on engaging subjects in the task affects eye saccades (following Ballard et al., 1995). Thus, it seems gravitational forces are background conditions in the first case, and eye saccades are (non-neural) components of the mechanism in the second (Kaplan, 2012).

Let's label those mechanisms underlying cognitive phenomena that span brain, body and world as 'extended mechanisms', and the position that extended mechanisms are responsible for at least some cognitive phenomena 'wide mechanism'.[10] Two views are subsumed by wide mechanism: one in which the bodily and environmental components of an extended mechanism count as 'cognitive', and a second in which bodily and environmental components do not necessarily count as cognitive, but the extended boundary of the mechanism is affirmed. The latter view results from supposing that components can contribute to a 'cognitive phenomena' (understood, say, as phenomena incorporating but not exhausted by cognition, or one studied by cognitive science) without themselves warranting the cognitive label. For instance, one might reserve 'cognition' to describe only those contributions that result from computational or neural processes, even when the total mechanism spans brain, body and world (for general discussion on issues around extended mechanisms, see Smart, 2022). What this possibility reveals is that views about what counts as 'cognitive' can come apart from views about the spread of a phenomena's constituents. Vernazzani (2019) adopts a position where this decouplability is apparent: "From the fact that a satisfactory explanation of the sensorimotor behavior of an agent must take into account mechanisms beyond the boundaries of the organism it does not follow that such mechanisms are cognitive in any relevant sense of the term" (p. 4549). Consider an analogy, borrowing from Kaplan (2012): the mechanism underlying a gecko's climbing ability appears to be distributed between organism and climbing substrate. The mechanism is thus 'wide'. The *biological* components of this mechanism, however, remain gecko bound.

These considerations show that two questions should not be equated when debating the appropriate unit of analysis: (i) do the constituents underlying phenomena studied by cognitive science include parts/processes in the organism's environment, and (ii) which of those mechanisms/processes count as 'cognitive'? I submit that enactivism is in tension with mechanistic explanations that restrict the constituents of perception, decision-making, social cognition and other cognitive phenomena to the brain or brain/body. However, as far as the 'unit of analysis' is concerned, enactivism is compatible with a version of wide mechanism that incorporates parts of the environment in the constituents of cognitive phenomena and considers these constituents as genuinely cognitive. Moreover, whilst there may be outstanding tension between enactivism and a version of wide mechanism that restricts cognition to

[10] Enactivists sometimes contrast their approach with one that focuses on 'internal mechanisms' (e.g., Gallagher, 2017, p. 6), without making a contrast to external mechanisms. Linking these notions is unhelpful to the degree it implies mechanistic explanation always appeals to internal parts and processes. As we have seen, this is not a given.

brain or body-bound components but which nonetheless allows for extended mechanisms underlying phenomena studied by cognitive science, there is at least *less* tension than with a version of mechanism that both restricts cognition to brain or body-bound components *and* denies extended mechanisms.[11] This is important because amalgamating issues (the spread of a phenomenon's constituents across body/world vs which constituents count as cognitive) can lead to an exaggeration of disagreement, for example, confusing the question of what our best model of cognition is with which parts of that model should be labelled 'cognitive'.[12]

At this stage, it is instructive to observe, following Herschbach (2012), that enactivists often respect the need to avoid something akin to the coupling-constitution fallacy (Adams & Aizawa, 2001, 2005), that is, the difference between constitutive and causal factors, however, subsequently conflate these on occasion. For instance, Herschbach observes that enactivist-leaning work on social cognition by De Jaegher et al., (2010) distinguishes between several roles for social interaction in an explanation of social cognition: contextual factors, enabling conditions, and constitutive elements (pp. 472–473). Indeed, enactivists often frame the concept of constitution in a standard fashion, i.e., in terms of a phenomenon's internal parts (albeit, in terms of operational closure or associated notions—see below), in contrast to its enabling factors. However, drawing on examples such as De Jaegher and Froese's (2009) discussion of social cognition, Herscbach observes that enactivists sometimes treat "any factor that is *necessary* for a cognitive process to be a *constituent* of that cognitive process, thus collapsing the distinction between enabling conditions and constitutive elements in the explanation of cognition" (p. 478). He also notes a tension latent in an enactivist distinction between cognisor and environment in terms of *autonomy*—roughly, the cognisor proper is an operationally closed system whose self-producing and self-maintaining parts are its constitutive elements, separate from the environment with which it interacts—and in terms of *phenomenological transparency*—roughly, the first-person experience of incorporating aspects of the environment into our bodies. Autonomy-based and transparency-based notions of constitution do not delineate the same boundaries.[13] Of course, one might seek to eliminate the distinction between coupling and constitution (for instance, in the manner of Ross & Ladyman, 2010; but for pushback see Kersten, 2016). In doing so though, one relinquishes the distinction between constitutive and enabling/contextual factors (or the

[11] For influential discussion around how to characterise cognitive phenomena that is critical of approaches such as enactivism, due to a wariness of conflating behaviour with cognition, see Aizawa (2015, 2017).

[12] These issues clearly bear on '4E' approaches, more broadly. However, for focus and brevity, my attention remains squarely on enactivism.

[13] Herschbach (2012) notes that defining cognition as 'relational', as enactivists sometimes do, won't get us out of the problem of demarcating the constituents of cognition because we can still ask what the constituents of the relational domain are: "Indeed, it seems what enactivists define as a social interaction is just such an autonomous relational domain consisting of two interacting agents" (p. 480).

like), that some enactivists may wish to preserve (Zahavi, 2003; Di Paolo, 2009; Thompson & Stapleton, 2009).[14]

In acknowledging that enactivists might sometimes confuse relations of dependence for evidence of constitution, it ought to be stressed that, at least in certain guises, enactivism has the resources to offer a relatively well-specified notion of constitution rooted in operational closure. According to this conception, constitution should be understood in terms of a closed network of parts/processes that depend on the network as a whole for their existence (see above). Thus, a cognitive system is comprised of parts/processes that realise the system as an adaptive autonomous system (in the first instance, an autopoietic system). This contrasts with constitution in terms of dependency relations. Thus, mechanists and enactivists appear concerned with different notions of constitution. Some comments are in order before we assume any troubling tension, however.

To the extent enactivists understand constitution in terms of precarious, closed networks, they are specifically concerned with autonomy which, at least when combined with adaptivity, forms the basis for cognition. By contrast, mutual manipulability is intended as a general criterion for what constitutes a mechanism underpinning any phenomenon. In turn, this has been leveraged (by Kaplan and others) to characterise the boundaries of cognition. However, there is nothing about mechanistic explanation that forces one to accept mutual manipulability (which has garnered a degree of controversy among even mechanists, e.g., see Leuridan, 2012). Moreover, even if one accepts mutual manipulability as evidencing the constitutive component parts/processes in a mechanism, one need not thereby accept that it satisfactorily demarcates cognition per se; in keeping with our presentation of wide mechanism, one might accept that the mechanism for some 'cognitive phenomena' includes some components which are not themselves, strictly speaking, *cognitive*.

Consider that mechanism alone does not offer any basis for characterising cognition. Rather, it provides a basis for understanding what constitutes any phenomenon. In turn, mutual manipulability provides an account of what evidences the constituents of a mechanism underlying a cognitive phenomenon but does not itself tell us what makes something cognitive. This opens space for enactivism to specify cognition, in terms of an adaptive-autonomous system, coupled with the world. Mechanists are not obliged to accept the enactivist conception of cognition, but they need not reject it either. In fact, mechanists might adopt enactivism as the theoretical starting point for characterising the phenomena that need explaining; enactivism identifies (cognitive) phenomena that may be subsequently investigated mechanistically (Lee & Millar, 2022). If enactivism provides the theoretical grounds for characterising cognitive phenomena—e.g., social cognitive phenomena as dynamic, multiple agent interactions—then mechanisms invoked to explain cognitive phenomena must overlap with the 'organism-environment totality'. Tensions may return, however, if (a) we accept that mechanistic investigations can 'reconstitute the phenomenon', suggesting mechanistic explanation retains ultimate authority in delineating

[14] A third possibility is to understand constitution as a special kind of causation but, following Kirchhoff (2017), this won't eliminate the operative distinction as there remains a difference between constitutive and non-constitutive causes.

the boundaries of phenomena, and (b) we think boundaries might realistically be redrawn such that the agent-world totality previously thought to underpin a quintessentially cognitive phenomenon is threatened. Such tensions underscore competing options for what takes precedence in characterising phenomena. We return to this in Sect. 4. For now, notice that one plausible position is to accept enactivism as providing guiding (but not immutable) heuristics for characterising phenomena subject to mechanistic investigation.

Another option not yet considered which eases friction between seemingly competing accounts of what comprises cognition involves accepting that there is more than one way to understand constitution. According to constitutional pluralism there are multiple, legitimate means of demarcating the constitutive boundaries of a cognitive system. This includes the mechanism-, autonomy-, and transparency-based criteria. If constitutional pluralism is correct, I suggest, the goal is not to establish which definition of constitution is correct but to appropriately individuate and avoid conflation. It also means that if mechanistic constitution does not capture everything of importance to enactivists, it does not follow that such constitution is at odds with enactivism.

3.2 Emergence & Downward Causation

Enactivists denounce reductionism (e.g., McGann et al., 2013, p. 203; Di Paolo & Thompson, 2014, p. 72). Given that mechanisms are defined in terms of organised parts, and the guiding heuristics of 'decomposition' and 'localisation' that affirm mechanistic explanation's targeting of a system's dissociable parts (e.g., Bechtel & Richardson, 2010), one might reasonably suspect that mechanism falls prey to what Thompson (2007) calls "part/whole reductionism", in which "all the properties of a whole are determined by the intrinsic (nonrelational) properties of its most fundamental parts". This contrasts with the holism of enactivism, according to which, "certain wholes possess emergent features that are not determined by the intrinsic properties of their most basic parts […] They are constituted by relations that are not exhaustively determined by or reducible to the intrinsic properties of the elements so related" (pp. 427–428). To support this holistic picture, enactivists appeal to the importance of emergence and downward causation.

Emergence is a slippery concept. However, in the present context, emergence broadly reflects the idea that a whole possesses causal powers that are non-reducible and non-identical to the causal powers possessed by its parts. In other words, emergence concerns whether higher levels have novel causal capacities or powers apart from their lower-level realisers. Downward causation concerns whether higher-level causes can have lower-level effects. These are related notions; if higher levels causally influence lower levels, then this indicates higher levels have causal powers that are not reducible to lower levels (cf. Paoletti & Orilia, 2017). Thus, emergence and downward causation overlap.

The key claim of this section is that mechanistic explanation is compatible with at least some plausible notions of emergence and downward causation. To the extent these interpretations do not undermine enactivism, mechanistic explanation is

compatible with enactivism. I will first paint a picture of levels in keeping with the mechanistic model that is non-reductive in character. This will then be used as the basis to draw more specific conclusions about the role of emergence and downward causation in mechanistic explanation.

To begin, notice that isolated components are not explanatory within a mechanistic model. The explanatory power of components lies in their organisational capacity to realise (collectively) a phenomenon. As such, the relations between often multitudinous (and often highly heterogeneous) components matter. As Bechtel sums up: "the working parts of a mechanism do different things than does the whole mechanism." (2008, p.146). Furthermore, mechanisms typically operate only under particular circumstances, in other words, their relationship with the environment matters too. Such considerations are reflected in the common refrain that a mechanism is more than the sum of its parts.

Mechanistic levels, moreover, need not correspond to an ontological hierarchy, in which the lowest level embodies the most ontologically fundamental stuff. First and foremost, levels are understood in organisational terms; a level refers to a set of identifiable working parts (Gillet, 2013). For example, in explaining blood circulation, we identify, roughly, the level of tissues within an organ, the level of communicating cells within a tissue, and the level of interacting organelles within a cell. Such levels are defined in terms of significant sets of entities and processes that bring about explanation-relevant effects. Organisational levels respect our intuitions about scale (higher levels correspond to a larger scale) whilst dispensing with fixed, meta-levels. From a mechanistic point of view, there are no explanation-independent ontological strata, and at the very least, mechanism is neutral about ultimate ontological reduction (cf. Oppenheim & Putnam, 1958; Churchland & Sejnowski, 1994; Wimsatt, 1994).

We can reinforce this non-reductive front by adopting a more partisan interpretation of mechanistic explanation. According to my preferred view, higher (organisational) levels are indeed constituted by lower levels (pace Craver & Bechtel, 2007) but this does not imply that higher levels are strictly identical to their parts (contra Fazekas & Kertesz, 2011). Here I follow in the footsteps of Piccinini (2020) who sets out an 'egalitarian ontology of levels', which averts the perils of reductionism whilst making clear why higher levels are explanatory. Key to this are the twin ideas that (1) wholes amount to *invariants* under transformations of their components (e.g., parts of my heart undergo many changes without my heart thereby disappearing), and (2) the properties of wholes (higher-level properties) are *aspects* of their realizers (lower-level properties), meaning they are not wholly distinct from their realisers (contra classical anti-reductionism) or identical to them (contra classical reductionism).[15] The central idea then is that levels are wholes that are subtractions from (or aspects of) their parts. By abstracting away from the parts, Piccinini suggests, we identify and track stable aspects of the world that are otherwise missed when we view all of a whole's parts. The point is as epistemic as it is metaphysical: "[E]ach level of description of a mechanism yields specific predictions that cannot

[15] This picture fits nicely with the adage that mechanistic explanation does not seek to *reduce* one level to another but *bridge* them (e.g., Ylikoski, 2012).

be made at other levels; because each level articulates essential information that is at best implicit at lower levels" (p. 316). A credible interpretation of mechanistic explanation thus institutes an egalitarian framework of levels that stands in contrast to reductionism.[16]

Using our non-reductive levels as a springboard, we can now dive into the implications for emergence. Povich & Craver (2018) provide a useful framing when they identify two senses of emergence that are unproblematic for mechanistic explanation:

1. Organisational emergence
2. Epistemic emergence

As we have seen, the mechanistic model allows that novel causal powers emerge from the organisation of interacting components (organisational emergence); mechanisms are frequently characterised by the capacities enabled by the spatial and temporal organisation of distinguishable parts, in contrast to aggregates (a distinction that can is detectable at least as far back as John Stuart Mill, 1843).[17] Moreover, we may not always be capable of decomposing a system due to cognitive limitations or resources (epistemic emergence). However, Povich & Craver (2018) also point to another and more problematic sense of emergence:

3. Ontic emergence

Ontic emergence here refers to phenomena with properties that have no constitutive explanation in terms of the organisation and activities of their parts. Mechanistic explanation does not allow for—or at least does not provide the resources to make sense of—such emergence. The activity of a whole mechanism, recall, is explained by showing how the constituent parts, organised as they are, collectively perform that same activity (although by adding a Piccininean twist, we can clarify that the higher-level phenomenon is only identical to an aspect of the part's activities, thus retaining the uniqueness, but not independence, of the higher level).

The types of emergence identified by Povich & Craver match a well-known distinction between 'diachronic' and 'synchronic' (or 'strong') emergence. It is enough for our purposes to note that diachronic emergence closely corresponds to organisation emergence. Synchronic emergence refers to the causal powers of a plurality of objects (a whole) in and above the causal powers of its collective parts and closely corresponds to ontic emergence. I join others in confessing my struggle to conceive of how wholes might possess causal powers that are absent in their parts, taken

[16] The issue of reduction raises a corresponding question about the role of abstraction versus detail in mechanistic explanation. Despite worries that mechanistic explanation implies that more detail is always better (Chirimuuta, 2014), mechanists have been at pains to stress the role of abstraction and idealisation in explanation (Miłkowski, 2016; Boone & Piccinini, 2016; Craver & Kaplan, 2020).

[17] Mechanism is arguably indifferent towards issues of reduction insofar as those concern fundamental 'being' (in a strong, metaphysical sense), and more preoccupied with the possibility of novel and empirically detectable patterns resulting from organisation, whatever the ultimate ontological nature of the entities involved might be (e.g., Winning & Bechtel, 2019).

collectively (for recent discussion, see Piccinini, 2020). Yet to the extent that enactivists can cash out emergence in terms of organisational and epistemic emergence, contention between enactivism and mechanistic explanation remains abated. We will return to what, exactly, the enactivist must say about emergence momentarily.[18]

What does the anti-reductionist picture of mechanistic levels suggest about downward causation? Following the interpretation established above, and what I take to be the prevailing position within mechanism, I suppose that causation for mechanists is 'intra level' (Craver & Bechtel, 2007); a higher level does not cause phenomena at lower levels because higher levels are *constituted* by lower levels (e.g., muscle cells comprise muscle tissue). Eronen (2013) helpfully identifies two things we might confuse when we talk about downward causation in mechanisms: "(1) causes that act from the mechanism as a whole to the components of the same mechanism, and (2) causation between entities of different (size) scales" (p. 1050). (1) is impossible, from a mechanistic perspective, "since composition is a form of non-causal dependency". (2) is unproblematic. Relations between wholes and parts are thus different from relations of scale (including space, time and force; cf. Eronen, 2013). It is likely, I suspect, that many intuitions about downward causation can be understood as tracking relations of scale. Indeed, some of the earliest uses of downward causation in a scientific context, such as Campbell's (1974) argument that downward causation occurs from higher to lower levels of biological organisation in natural selection, can be interpreted through relations of scale.[19]

Enactivists might worry, at this point, that we have not done justice to their view of part-whole relations. Of particular note, Thompson (2007) elucidates an enactivist conception of emergence and downward causation in terms of 'dynamic co-emergence', which refers to the way wholes and parts define each other. Drawing on Kronz and Tiehen (2002), Thompson summarises that "Dynamic co-emergence means that part and whole co-emerge and mutually specify each other" (p. 431). Acknowledging the importance of dynamic co-emergence for enactivism, I propose three kinds of relations that help to specify how wholes and parts may 'co-emerge' and 'mutually specify' one another; I claim that none of these prevents the enactivist from embracing mechanistic explanation.[20]

The first concerns 'dialectical relations' whereby the individuation or categorisation of a component part/process is determined, in part, by the whole in which it participates. Such relations are straightforwardly recognised within orthodox mechanism to the extent that wholes are constituted by components whilst components are defined by the phenomenon they realise; there are no parts without the phenomenon

[18] Some may harbour a lingering suspicion that mechanisms cannot capture the open-ended, cyclic nature of cognitive processes. Concepts such as 'maintenance mechanisms', introduced above, capture such processes, and there is nothing contradictory about mechanistic explanation and the repeated, open-ended or continuous nature of a phenomenon. Even explanations that emphasise a product can allow the process to be continuous, with the product itself potentially feeding-back. Homeostatic mechanisms for the ongoing production of stable states, for instance, are still mechanisms.

[19] Eronen (2013) in fact recommends we discard talk of entities being at the same or different 'levels', and only talk about different scales.

[20] I would like to thank an anonymous referee for pressing me on this section and assisting in the elucidation of the three types of relations.

they collectively cause, produce or maintain (e.g., Glennan, 2002). This makes sense of the fact that one and the same system, e.g., an electrical component, can function in different ways depending on the circuit in which it is embedded, e.g., a motor or generator. Moreover, one and the same mechanism (particularly biological mechanisms) frequently play multiple roles relative to different phenomena they help realise, e.g., kidneys regulate blood pressure and detoxify the blood.

The second concerns 'enabling relations' whereby a component part/process gains new capacities in virtue of the whole it participates in (mapping closely to the concept of 'organisational emergence'). Consider how, say, a plant cell is energetically sustained by being part of a leaf that is structured in such a way as to maximise light absorption (which is supported by being part of a larger leaf-and-branch structure and so on). By being organised within a leaf, a plant cell is restricted in novel ways (e.g., limited cell migration) but can partake in activities it could not in isolation. A mechanistic perspective should thus grant that parts face "context-free constraints" in a manner Thompson suggests (for an extended discussion on 'causes as constraints', see Juarrero, 1999). Mechanists may unproblematically affirm the enactivist notion that constraints in general "can be understood as relational properties that the parts possess in virtue of their being integrated or unified (not aggregated) into a systemic network" (p. 424) and that a context-sensitive constraint is "one that synchronizes and correlates previously independent parts into a systematic whole" (p. 425).

The third concerns 'existential relations' whereby component parts/processes are not only constrained by and gain new powers in virtue of being part of a whole they help realise but depend on the whole for their existence. The importance of these relations for the enactivist conception of cognition, rooted in adaptive autonomy, is captured in the notion of operational closure. As already noted, processes in living cells, such as metabolic processes, existentially depend on being part of a closed network of other mutually dependent parts and processes, such as being contained within a membrane that itself is sustained by metabolic processes. Such existential relations help demarcate living systems from machines. For instance, whilst an artificial electrical component bears dialectical relations (e.g., it is individuated as a motor or generator, say, depending on the circuit it is part of) and enabling relations (e.g., the capacity to convert electrical energy into kinetic energy or vice versa, say, depends on the circuit it part of), the component does not existentially depend on being part of a larger mechanism. However, not all mechanisms are machines and much ink has been spilt by mechanists to distance their model from more 'classical' mechanistic explanation associated with the machine analogy. For instance, as Levy and Bechtel (2020) argue, biological mechanisms are unlike classical machines in often possessing 'concentrations' rather than 'discrete parts', and dynamically changing their constitution over time. Thus, explaining cognition mechanistically does not imply cognitive systems are akin to classical machines and does not threaten the importance of existential relations for understanding life (and thus cognition) according to enactivists.

As an aside, an enactivist sceptic might complain that once distanced from machines, mechanistic explanation begins to look trivial: was it not the comparison between living systems and artificial machines that made classical mechanism

significant? Whatever one thinks of the historical relation between new and classical mechanism and the merits of each, however, new mechanism seems far from trivial, generating rich internal debates among proponents on a range of issues (e.g., Glennan & Illari, 2018), as well as between proponents and detractors who take new mechanism to be interesting enough to be inadequate (e.g., Chirimuuta, 2014).

To close our argument for the compatibility between mechanism and enactivism on emergence and downward causation, we can turn to a passage that draws attention to an important sense of unity between levels for enactivism. Thompson (2007) invites caution about the "downward" metaphor when writing:

> It is questionable whether this metaphor is a good one. Although there are clearly empirical differences in scale and logical differences in order between the topology of a system and its constituents processes and elements, the two levels do not move in parallel, with one acting upward and the other acting downward, because the whole system moves at once. (p. 426).

Thompson's open-minded attitude toward the causal status of downward causation sits well with a certain mechanistic point of view. To repeat, higher levels do not cause things to happen at lower levels because higher levels are (aspects of) lower levels—everything "moves at once". However, the properties of components are affected by being members of an organised set of interdependent elements. In summary: things at a larger scale cause things to happen at a smaller scale, and parts may gain and lose powers by participating in a whole mechanism. To the extent these senses of 'downward causation' comply with enactivism, there is less tension between mechanistic explanation and enactivism than may first appear.

3.3 Dynamical Descriptions & Reciprocal Causation

In stressing the necessity of agent-environment interaction for cognition, enactivists underscore the role of continuous reciprocal causation, whereby two or more systems are coupled such that each's behaviour determines the other in a rolling cycle of influence. As we saw already, for many enactivists, the interacting organism-environment totality *is* the unit of analysis; cognition is "intrinsically relational and dynamic in nature" (McGann et al., 2013, p. 230). Continuous reciprocal causation has also been thought to undermine the possibility of decomposition and localisation, the guiding heuristics of mechanistic explanation (e.g., Van Gelder, 1995). This explains the tendency of enactivists to appeal to dynamical models, and at the same time, drive a wedge between dynamical descriptions and mechanistic explanations.

Dynamical systems theory (DST) is a framework for modelling and describing systems that change over time. Crucial to DST are well-established geometric concepts, such as a state-space or phase space (the set of all possible states that a system can assume over time), which can be analysed in terms of its control parameters and collective variables. The system's potential trajectories can be understood by combining different values with the relevant equation. It provides a set of mathematical tools for describing the behaviour of a complex system that evolves over time,

typically involving differential or difference equations. DST thus plays an important part in describing the time-dependent processes in neural systems, from single neurons to whole-brain networks. In turn, DST is well-placed to describe relations of interaction and continuous reciprocal causation (e.g., Clark, 2014, p. 154). To borrow from Chemero & Silberstein: "dynamical systems theory is especially appropriate for explaining cognition as interaction with the environment because single dynamical systems can have parameters on each side of the skin" (Chemero & Silberstein, 2008, p. 14). Unsurprisingly, enactivists often adopt DST as an ally in combating cognitivist approaches that centre the activities of individuals and their brains as the explanans of cognitive phenomena.

DST has been the target of debate concerning the nature of explanations in cognitive science. One popular idea, which we will dub 'dynamism', holds that DST reflects a fundamentally different form of explanation from mechanistic explanation, with its focus on mathematical description and (apparent) autonomy from the identification of decomposable components, their operations, and their organisation (e.g., Chemero & Silberstein, 2008; Silberstein & Chemero, 2013; Stepp et al, 2011). For example, Stepp et al (2011) write,

> Dynamical explanations do not propose a causal mechanism that is shown to produce the phenomenon in question. Rather, they show the change over time in a set of magnitudes in the world can be captured by a set of differential equations. (p. 432)

If enactivists explain exclusively by appealing to DST, and DST is independent of mechanistic explanation, then enactivists do not explain by appealing to mechanisms. This firm separation of dynamical and mechanistic explanations coincides with an averred opposition between dynamical and computational explanations, which enactivists also typically reject (e.g., Varela, Thompson & Rosch, 1991/2017; cf. van Gelder, 1995). This is because DST offers an alternative to computational explanations that are themselves mechanistic in nature, the thought goes (Zednik, 2011). To clarify, dynamists are not committed to claiming that mechanisms are never explanatory. Rather, dynamists are committed to claiming that dynamical explanation is non-mechanistic because variables in a dynamical description do not refer to mechanisms or their activities (for discussion on the prospects of explanatory pluralism, see Dale, 2008; Dale et al., 2009).

The main claim of this section is that dynamical and mechanistic explanations can be integrated, thus further easing the tension between enactivism and mechanistic explanations. It has already been argued at length that dynamical and mechanistic explanations are compatible (e.g., Bechtel & Abrahamsen, 2010; Zednik, 2011; Kaplan, 2015, 2017; Miłkowski et al., 2018). To sidestep familiar ground, I will outline what I take to be the main moves available to those seeking reconciliation, along the way, noting novel considerations as they pertain to our task of evaluating the relationship between enactivism and mechanism.

There are multiple ways of construing the type of explanation DST offers (cf. Kaplan, 2018). For present purposes, I will focus on the debate between 'covering law' interpretations—dynamical models explain because they are a special case of

covering law explanations—and 'mechanistic' interpretations—dynamical models are explanatory because they describe mechanisms. There are negative and positive arguments supporting the mechanistic interpretation and thus the integrability of dynamic and mechanistic explanations (and thus the compatibility of mechanistic and enactivism). Negative arguments claim that dynamical explanations descriptions cannot be explanatory unless integrated with mechanistic explanation either because of inherent flaws in the covering-law model (for some influential criticism, see Eberle, Kaplan, & Montague, 1961; Forge, 1980; Salmon, 1984) or because of difficulties interpreting dynamical descriptions in terms of laws (e.g., Kaplan, 2018), i.e., the covering law interpretation is false. Positive arguments claim that dynamical explanations do, as a matter of fact, feature in or alongside mechanistic explanations, and this proves at least some DST-type explanations are interconnected with mechanistic ones. As the foibles of the covering-law model have been discussed at greater length, we will focus on the positive arguments.

Perhaps the most common positive claim of those defending the integration of mechanistic and dynamical explanations is that the latter describes and thus presupposes mechanisms, or they only explain when they do so. Biologically plausible mechanisms are dynamic systems: brains are complex systems that evolve over time, for instance. A further step is then required to claim that dynamical descriptions explain *because* they describe mechanisms. Providing some support for this notion is the trivial fact that, as Zednik (2011) notes, among other dynamical tools, differential equations describe spatial and temporal relations between structural and functional properties of mechanisms and their evolution over time.

Also affirming the role of dynamical descriptions in mechanistic explanation, whilst acknowledging its neglect among theorists, Kaplan (2015) writes that, "dynamics have always had a proper place in the mechanistic framework under the guise of temporal organization, although its role in mechanistic explanations has been seriously underemphasized" (p. 773). Bechtel & Abrahamsen make a similar point, arguing that the temporal 'orchestration' of components has been undervalued in the mechanism literature (such as the role of negative and positive feedback, and self-organization). To compensate, they offer a revised version of their classic definition of a mechanism (quoted above):

> A mechanism is a structure performing a function in virtue of its component parts, component operations, and their organization. The orchestrated functioning of the mechanism, manifested in patterns of change over time in properties of its parts and operations**,** is responsible for one or more phenomena. (p. 323)

In short, the idea is that dynamical descriptions play a role in describing the temporal organisation and evolution of parts and processes within mechanisms.

There is some room for differences of opinion within the mechanistic interpretation, distinguishing 'strong mechanism'—roughly, the claim that mechanistic explanation is the exclusive kind of explanation and DST can only explain to the extent it can assimilate into mechanistic explanation—and 'moderate mechanism'—roughly, the claim that dynamical descriptions explain, at least sometimes and at least in part, due to their intersection with mechanistic explanation. If strong mechanism is right,

then enactivism must be compatible with mechanistic explanation in virtue of offering dynamical explanations. But moderate mechanism is sufficient for dynamical and mechanistic explanation to be compatible, and thus ease the tension between enactivism and the mechanistic model. In any case, to be maximally concessional, I wish only to defend moderate mechanism.

Zednik (2011) captures the spirit of moderate mechanism when he writes,

> The general point is this: like English, the mathematics of dynamical systems theory is a language, an important feature of which is its capacity to represent. What is being represented—a mechanism, law of nature, or my neighbor's pet iguana—is not determined by the language being used but by the way in which tokens of that language are interpreted. (p. 247).

He further clarifies that some (but not all) canonical dynamical models clearly describe mechanism components and their organisation (drawing on canonical examples such as Thelen et al., 2001, dynamical field theory model of infant perseverative reaching). DST may indeed be uniquely well-placed to describe mechanisms with parts located across brain, body, and the environment and which are engaged in 'continuous reciprocal causation'. In other words, the sorts of mechanisms that likely play a role in cognition.

A somewhat different but overlapping take on 'moderate mechanism' that I endorse is adopted by Lyre (2018), who argues that dynamical and mechanistic approaches do offer different explanatory 'perspectives': horizontal (dynamical) and vertical (mechanistic). He thus rejects both what he terms 'strong dynamism'—the view that dynamical explanations are self-contained and have nothing to do with mechanisms—as well as 'strong mechanism'—the view that mechanistic models exhaust explanation. Nonetheless, these perspectives intersect: "[D]ynamical explanations possess underlying mechanisms not only as realizers, but also as "intersection points" of the horizontal and vertical direction of explanation" (p. 5153). What matters for our purposes is the idea that dynamical equations track "spatiotemporal-cum-causal relational properties of a dynamical system" and this corresponds to the *organisational structure* of an underlying mechanism that realises a phenomenon (p. 5142).

Following Lyre (2018), I contend that a crucial characteristic of dynamical explanations is their capacity to describe 'higher level' properties shared by otherwise different realisers, for example, the oscillator equation that describes the behaviour of many artificial and neural systems. Dynamical explanations are 'structurally grounded', meaning "they individuate their entities only relationally by focusing on the relevant spatiotemporal-cum-causal structure of their target systems" (p. 5147). Importantly, mechanisms realise instances of dynamical laws. In turn, dynamical variables correspond to components and their organisation, and dependencies among dynamic variables correspond to causal relations between components. 'Higher-level structure' is not something independent of 'lower-level structure' but refers to one and the same set of organised components and their relations. What dynamical descriptions offer is quantification over causally relevant relations, "of the whole class of realizing mechanism that fall under their scope" (p. 5152). In my

preferred terms, dynamical explanations track patterns of organisational structure among a class of realising mechanisms.

In closing this section, I want to highlight one area where enactivists appear systematically to evoke mechanistic explanations. Consider the 'sensorimotor laws' studied by sensorimotor theory—the enactivism offshoot introduced above that is typically encompassed by autopoietic enactivism. On the face of it, appealing to 'laws' with little mention of causally relevant parts and processes suggests a type of covering law explanation—one perhaps divorced from mechanisms after all. Following Vernazzani (2019), however, a decompositional strategy seems implicit in O'Regan and Noë's (2001) canonical outline of sensorimotor theory, insofar as they point to *distinct* regularities related to *different* features of visual perception, such that sensorimotor theory implicates localisable components that are responsible for distinct sets of sensorimotor operations. Vernazzani observes that SME allows for "Distinct sensorimotor contingencies for specific characteristics" and "Different neural structures related to different characteristics" (p. 4546). This suggests a strategy of decomposition and localisation. Concurrently, sensorimotor theory provides few details about the avouched neural structures. SME may thus be seen to offer blueprints for 'mechanism sketches'. Such sketches refer to outlines of mechanisms that identify functional properties whilst omitting all or many structural details that are ultimately required for complete understanding (Machamer et al., 2000; Piccinini & Craver, 2011). We will touch on this idea again in Sect. 4.

To summarise: dynamical descriptions are at least sometimes explanatory because they describe mechanisms or, at the very least, complement mechanistic explanations by revealing higher-level structures shared by mechanisms. Therefore, enactivism's preoccupation with dynamical descriptions need not imply a lack of concern for underlying mechanisms. Moreover, despite appearing to offer covering law explanations, we saw evidence that sensorimotor theory delivers mechanism sketches and is thus congruent with mechanistic explanation.

4 The Value of Dialogue

We have so far laboured to close the gap between enactivism and mechanism. Mechanistic explanation is neither narrow nor reductive, can incorporate elements of body and world (either as constitutive or causally necessary factors), is compatible with a plausible version of emergence, and intersects with the dynamical descriptions that make sense of complex causal reciprocity. If correct, mechanistic explanation is less at odds with enactivism than may first appear. Nevertheless, we have said little about the value of dialogue between enactivism and mechanism. In this section, I gesture toward how enactivists benefit from embracing mechanistic explanation, what practical value those conducting mechanistic modelling might gain from attention to enactivism, as well as some outstanding tensions.

The broad value of mechanistic explanation for enactivism is straightforward: if cognitive science profits from mechanistic explanation, then the congruence of enactivism and mechanistic explanation indicates some compatibility between enactivism and practising science (for an indication of what 'enactive

mechanistic explanation' might look like, albeit in relation to radical enactivism, see Abramova & Slors, 2019). More narrowly, and more interestingly, mechanism may ensure empirical content for certain enactivist claims, such as those concerning cognitive boundaries. Evaluating whether cognition incorporates extra-bodily constituent is notoriously hard, given perennial disagreement over how to evaluate the material/temporal scope of cognition. Outstanding issues regarding the possible decoupling of 'cognitive phenomena' from those constituents which count as truly *cognitive* aside (see Sect. 3.1), mechanism provides one set of relatively clear standards: cognition is constitutively wide when the mechanisms for cognition are wide, understood in terms of mutual manipulability. This exposes cognitive boundaries to empirical scrutiny (at least in principle). In Kaplan's (2012) words, mechanism provides the resources to turn a claim about the extendedness of cognition from intuition-based speculation or purely function-based comparisons (Clark & Chalmers, 1998) into a "legitimate empirical hypothesis amenable to experimental test and confirmation" (2012, p. 546).

Such dependency on the contingent discoveries of disciplines like cognitive neuroscience may unsettle some enactivists, and I believe this is where residual tension may lie between enactivism and what we might call a 'mechanisms-first' approach (cf. Lee & Millar, 2022). Proponents of the mechanistic model, as we saw, underscore the possibility of mechanistic investigation 'reconstituting' the phenomenon to be explained as details emerge of how causal parts/processes map to effects, implying that mechanistic investigation has the authority to specify the nature of a cognitive phenomenon based on its contingent results. The extendedness of cognition may be less negotiable for the enactivist convinced by the essential agent-environment nature of mentality. Nevertheless, there are two ways in which tension here may be illusory. First, enactivism may be construed as placing an empirical bet; the point is less that cognition is constituted by the agent-environment unit by some unquestionable theoretical fiat and more that such constitution offers our best starting conception of cognitive phenomena given a balance of theoretical considerations and empirical evidence (Lee & Millar, 2022). Taken this way, the enactivist may welcome mechanistic scrutiny. Second, enactivists may simply have different but compatible standards of constitution: one might grant that *as per the mechanist's sense of constitution*, whether the mind is partially constituted by the environment is indeed determined empirically via the uncovering of mechanisms, whilst in another sense of primary importance, cognition necessarily incorporates agent-environment interaction. This invites the possibility of multiple (non-mutually exclusive) senses of constitution—in other words, the constitutional pluralism gestured at above (Sect. 3.1),

Our efforts so far have focused on easing the tension between enactivism and mechanism by elucidating how the latter does not entail narrow and reductive explanations. We have not, however, addressed the worth of enactivism for those engaged in mechanistic modelling of cognition. For instance, we might ask what practical benefits enactivism affords the cognitive neuroscientist engaged in uncovering the neural mechanisms of perception, decision-making or learning. Broadly, I propose that enactivism's value for those engaged in mechanistic modelling, especially in cognitive neuroscience, is primarily 'corrective' and 'heuristic'. By corrective I

mean enactivism challenges neuro-centrism and explanations that discount organism-environment interaction. By heuristic I mean it draws our attention to certain configurations of processes and components, that is, which parts and activities of the organism/world may underlie cognitive phenomena. Enactivism especially asks us to consider, on the negative side, how a cognitive phenomenon might be realised without computation or representation, and on the positive side, what role continuous and reciprocal interaction with the environment might play. Thus, enactivism draws our attention to potential (extended) mechanisms and their environmental modulation. One concrete output of this role is the offering of mechanism sketches, of the sort we saw were suggested by sensorimotor theory (Sect. 3.3).

A similar perspective is adopted by Miłkowski et al. (2018). The authors worry that 'wide approaches', including enactivism, often make grand claims such as "embodiment is essential to cognition" or "cognitive phenomena are always constituted by interactions with the environment". These claims suffer from remaining "fairly abstract and focused on deciding yes–no questions rather than building unified models of cognitive phenomena" (p. 2). This generates a corresponding concern that "Grand issues in the study of cognition cannot be fruitfully understood in terms of a series of simple dichotomies" (p. 2). Miłkowski et al. (2018) indicate that wide approaches like enactivism are akin to 'grand research traditions' such as computationalism. Enactivism and computationalism are alike in largely failing to generate novel predictions or explanations for particular cognitive phenomena. Their value lies, instead, in providing 'guiding heuristics'. Using the case of group decision-making, they write:

> A proponent of traditional computational modeling would ask what the overall task is and why solving it is appropriate; what the algorithms and representations involved are; and how they are physically implemented […] the enactive perspective (at least in its non-classical version) points to participatory negotiation of how the activity is perceived by various agents involved, and the distributed perspective hints that the phenomenon may involve not only human agents but also external representations and instruments. (p. 4)

This passage demonstrates how a mechanisms-first approach might accord mechanistic modelling priority in determining the nature of a cognitive phenomenon. As Miłkowski et al. (2018) note, if we treat wide approaches as offering 'generic heuristic advice' then they are not the final arbiters of what parts of the world are responsible for which phenomena and where those are located. For example, whilst wide approaches, such as an enactivism, constructively draw our attention to non-computational/non-representational resources, and rightly demand that we find "additional causal evidence from lower levels of mechanistic organization in order to talk of computation and representation" (p. 13), they are not conclusive for determining the ultimate explanatory status of computation and representation.

Such considerations of computation and representation bring us to another area of unresolved tension. On the face of it, any compatibility between enactivism and mechanistic explanation points to a congruence between enactivism and contemporary cognitive neuroscience, insofar as the latter prioritises dynamic mechanistic explanations. However, explanations in cognitive neuroscience have been

characterised in terms of (dynamical and embodied) computational mechanisms performing operations over representations (e.g., Piccinini, 2020, 2022). Enactivism is deeply suspicious, at best, and wholly hostile, at worst, toward explanations in terms of computation/representation. A couple of responses are available.

First, one might bite the bullet and acknowledge the consistency of enactivism and mechanistic explanations of cognition *insofar as* the latter do not involve computation or representation. The cost of such tenacity is the burden of showing either that the characterisation of cognitive neuroscience in terms of computational mechanisms is wrong, or that current cognitive neuroscience itself is misguided. Second, one might use the more general compatibility between mechanistic explanation and enactivism, explored in this paper, as an impetus to investigate whether computation and/or representation in cognitive neuroscience (much revised since the days of classical cognitive science), is more amenable to enactivism than previous iterations. Within this territory lie further options. For instance, one might exploit attempts to quarantine computation from representation—by appealing to mechanism—to argue that enactivism is compatible with computation so long as it remains non-semantic (Villalobos & Dewhurst, 2017). A more reconciliatory method might use recent reforms of the concept, facilitated by developments in cognitive neuroscience, to argue that enactivism need not reject representation after all. Exploring this space of possibilities will be the task for another day.[21]

5 Conclusion

Dichotomies force us to choose one possibility at the expense of another, losing any potential the discarded option may have afforded. To avoid the needless abandonment of valuable perspectives in philosophy of cognitive science, we must be careful to separate real dichotomies from illusory ones borne from our polarising tendencies. This paper attempted to weaken apparent tensions between enactivism and mechanistic explanation.

At first glance, enactivism favours wide and non-reductive explanations whilst mechanistic explanations are narrow and reductive. However, new mechanism can be construed such that enactivists can accept or develop mechanistic explanations without compromising their core tenets. At the same time, enactivism offers corrective and heuristic value for those engaging in mechanistic modelling. Nevertheless, a disparity in starting point between enactivism and mechanism may endure if the latter lends itself to viewing the constitutive elements of cognition as being settled by cognitive neuroscience and similar disciplines engaged in mechanistic modelling of cognitive phenomena. Somewhat addressing these concerns, we saw the possibility of further easing strain by employing a sufficiently pluralistic conception of constitution. Disagreement between enactivism and cognitive neuroscience may also arise if the latter is construed

[21] Various contemporary accounts of representation eschew classical language-like symbols bearing purely declarative content. Moreover, recent attempts to naturalise content-determining relations have turned to the contribution of representation-like mechanisms to the self-organised, goal-directed behaviours of organisms, reminiscent of enactivist approaches to naturalising teleology (e.g., Lee, 2021; Piccinini, 2022). Nonetheless, some enactivists, especially in the 'radical' branch, have clarified their hostility toward the value of subpersonal representation, however conceived (e.g., Hutto & Myin, 2012).

as offering explanations in the form of computational and/or representational mechanisms, and the former is unable to tolerate such practices. Future research should attend to the relationship between enactivism and computation/representation as they appear in contemporary cognitive neuroscience, exploring the possibility that the translation of such controversial notions into mechanistic vocabulary may blunt the edge of hostility.

Acknowledgements The author would like to thank Paco Calvo, Becky Millar and Joe Dewhurst for their helpful comments on an earlier draft.

Funding Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature. This research was supported by a Juan de la Cierva Fellowship from Ministerio de Ciencia e Innovación del Gobierno de España (Award # FJC2019-041071-I).

Declarations

Conflict of interest The author has no financial or non-financial interests to declare.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Abramova, E., & Slors, M. (2019). Mechanistic explanation for enactive sociality. *Phenomenology and the Cognitive Sciences, 18*(2), 401–424. https://doi.org/10.1007/s11097-018-9577-8

Adams, F., & Aizawa, K. (2001). The bounds of cognition. *Philosophical Psychology, 14*(1), 43–64. https://doi.org/10.1080/09515080120033571

Aizawa, K. (2015). What is this cognition that is supposed to be embodied? *Philosophical Psychology, 28*(6), 755–775. https://doi.org/10.1080/09515089.2013.875280

Aizawa, K. (2017). Cognition and behavior. *Synthese, 194*(11), 4269–4288. https://doi.org/10.1007/s11229-014-0645-5

Aizawa, K., & Adams, F. (2005). Defending non-derived content. *Philosophical Psychology, 18*(6), 661–669. https://doi.org/10.1080/09515080500355186

Ballard, D. H., Hayhoe, M. M., & Pelz, J. B. (1995). Memory representations in natural tasks. *Journal of Cognitive Neuroscience, 7*(1), 66–80. https://doi.org/10.1162/jocn.1995.7.1.66

Barandiaran, X. E. (2017). Autonomy and enactivism: Towards a theory of sensorimotor autonomous agency. *Topoi, 36*(3), 409–430. https://doi.org/10.1007/s11245-016-9365-4

Bechtel, W. (1998). Representations and cognitive explanations: Assessing the dynamicist's challenge in cognitive science. *Cognitive Science, 22*(3), 295–318. https://doi.org/10.1207/s15516709cog2203_2

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

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. https://doi.org/10.1016/j.shpsa.2010.07.003

Bechtel, W., & Richardson, R. C. (1993). *Discovering complexity: Decomposition and localization as strategies in scientific research*. Princeton University Press.
Bechtel, W., & Richardson, R. C. (2010). *Discovering complexity: Decomposition and localization as strategies in scientific research*. MIT press.
Bich, L., & Arnellos, A. (2012). Autopoiesis, autonomy, and organizational biology: Critical remarks on 'Life after Ashby'. *Cybernetics & Human Knowing, 19*(4), 75–103.
Boone, W., & Piccinini, G. (2016). Mechanistic abstraction. *Philosophy of Science, 83*(5), 686–697. https://doi.org/10.1086/687855
Campbell, D. T. (1974). 'Downward causation' in hierarchically organised biological systems. *Studies in the philosophy of biology* (pp. 179–186). Palgrave.
Chemero, A., & Silberstein, M. (2008). After the philosophy of Mind: Replacing scholasticism with science. *Philosophy of Science, 75*(1), 1–27. https://doi.org/10.1086/587820
Chirimuuta, M. (2014). Minimal models and canonical neural computations: The distinctness of computational explanation in neuroscience. *Synthese, 191*(2), 127–153. https://doi.org/10.1007/s11229-013-0369-y
Churchland, P. S., & Sejnowski, T. J. (1994). *The computational brain*. MIT Press.
Clark, A. (2014). *Mindware: An introduction to the philosophy of cognitive science* (2nd ed.). Oxford University Press.
Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis, 58*(1), 7–19.
Craver, C. F. (2007). *Explaining the brain: Mechanisms and the mosaic unity of neuroscience*. Oxford University Press.
Craver, C. F., & Bechtel, W. (2007). Top-down causation without top-down causes. *Biology & Philosophy, 22*(4), 547–563. https://doi.org/10.1007/s10539-006-9028-8
Craver, C. F., & Kaplan, D. M. (2020). Are more details better? On the norms of completeness for mechanistic explanations. *The British Journal for the Philosophy of Science, 71*(1), 287–319. https://doi.org/10.1093/bjps/axy015
Currie, G. (2001). Methodological individualism: Philosophical aspects. *International Encyclopedia of the Social & Behavioral Sciences*. https://doi.org/10.1016/B0-08-043076-7/01028-7
Dale, R. (2008). The possibility of a pluralist cognitive science. *Journal of Experimental & Theoretical Artificial Intelligence, 20*(3), 155–179. https://doi.org/10.1080/09528130802319078
Dale, R., Dietrich, E., & Chemero, A. (2009). Explanatory pluralism in cognitive science. *Cognitive Science, 33*(5), 739–742. https://doi.org/10.1111/j.1551-6709.2009.01042.x
Eberle, R., Kaplan, D., & Montague, R. (1961). Hempel and Oppenheim on explanation. *Philosophy of Science, 28*(4), 418–428.
De Jaegher, H., & Di Paolo, E. (2007). Participatory sense-making: An enactive approach to social cognition. *Phenomenology and the Cognitive Sciences, 6*(4), 485–507. https://doi.org/10.1007/s11097-007-9076-9
De Jaegher, H., Di Paolo, E., & Gallagher, S. (2010). Can social interaction constitute social cognition? *Trends in Cognitive Sciences, 14*(10), 441–447. https://doi.org/10.1016/j.tics.2010.06.009
De Jaegher, H., & Froese, T. (2009). On the role of social interaction in individual agency. *Adaptive Behavior, 17*(5), 444–460. https://doi.org/10.1177/1059712309343822
Forge, J. (1980). The structure of physical explanation. *Philosophy of Science, 47*(2), 203–226.
De Jesus, P. (2016). Autopoietic enactivism, phenomenology and the deep continuity between life and mind. *Phenomenology and the Cognitive Sciences, 15*(2), 265–289. https://doi.org/10.1007/s11097-015-9414-2
Di Paolo, E. A. (2005). Autopoiesis, adaptivity, teleology, agency. *Phenomenology and the Cognitive Sciences, 4*(4), 429–452. https://doi.org/10.1007/s11097-005-9002-y
Di Paolo, E. (2009). Extended life. *Topoi, 28*(1), 9. https://doi.org/10.1007/s11245-008-9042-3
Di Paolo, E., Buhrmann, T., & Barandiaran, X. E. (2017). *Sensorimotor Life: An enactive proposal*. Oxford University Press.
Di Paolo, E., & Thompson, E. (2014). The enactive approach. In L. Shapiro (Ed.), *The Routledge handbook of embodied cognition* (pp. 68–78). Routledge.
Eronen, M. I. (2013). No levels, no problems: Downward causation in neuroscience. *Philosophy of Science, 80*(5), 1042–1052. https://doi.org/10.1086/673898
Fazekas, P., & Kertész, G. (2011). Causation at different levels: Tracking the commitments of mechanistic explanations. *Biology & Philosophy, 26*(3), 365–383. https://doi.org/10.1007/s10539-011-9247-5
Fisk, J., Lackner, J. R., & DiZio, P. (1993). Gravitoinertial force level influences arm movement control. *Journal of Neurophysiology, 69*(2), 504–511. https://doi.org/10.1152/jn.1993.69.2.504

Fodor, J. (1980). Methodological solipsism considered as a research strategy in cognitive psychology. *Behavioral and Brain Sciences, 3*(1), 63–73. https://doi.org/10.1017/S0140525X00001771
Fodor, J. (1987). *Psychosemantics: The problem of meaning in the philosophy of mind*. MIT Press.
Froese, T., & Di Paolo, E. A. (2011). The enactive approach: Theoretical sketches from cell to society. *Pragmatics & Cognition, 19*(1), 1–36. https://doi.org/10.1075/pc.19.1.01fro
Fuchs, T. (2017). *Ecology of the brain: The phenomenology and biology of the embodied mind*. Oxford University Press.
Gallagher, S. (2017). *Enactivist interventions: Rethinking the mind*. Oxford University Press.
Gillett, C. (2013). Constitution, and multiple constitution, in the sciences: Using the neuron to construct a starting framework. *Minds and Machines, 23*(3), 309–337. https://doi.org/10.1007/s11023-013-9311-9
Glennan, S. (2002). Rethinking mechanistic explanation. *Philosophy of Science, 69*(S3), S342–S353.
Glennan, S., & Illari, P. M. (Eds.). (2018). *The Routledge handbook of mechanisms and mechanical philosophy*. Routledge.
Hedström, P., & Ylikoski, P. (2010). Causal mechanisms in the social sciences. *Annual Review of Sociology., 36*, 49–67. https://doi.org/10.1146/annurev.soc.012809.102632
Hempel, C. G. (1965). *Aspects of scientific explanation and other essays in the philosophy of science*. New York: Free Press.
Herschbach, M. (2012). On the role of social interaction in social cognition: A mechanistic alternative to enactivism. *Phenomenology and the Cognitive Sciences, 11*(4), 467–486. https://doi.org/10.1007/s11097-011-9209-z
Hutto, D. D., & Myin, E. (2012). *Radicalizing enactivism: Basic minds without content*. MIT Press.
Juarrero, A. (1999). *Dynamics in action: Intentional behavior as a complex system*. MIT Press.
Kaiser, M. I., & Krickel, B. (2017). The metaphysics of constitutive mechanistic phenomena. *British Journal for Philosophy of Science, 68*(3), 745–779. https://doi.org/10.1093/bjps/axv058
Kaplan, D. M. (2012). How to demarcate the boundaries of cognition. *Biology & Philosophy, 27*(4), 545–570. https://doi.org/10.1007/s10539-012-9308-4
Kaplan, D. M. (2015). Moving parts: The natural alliance between dynamical and mechanistic modeling approaches. *Biology & Philosophy, 30*(6), 757–786. https://doi.org/10.1007/s10539-015-9499-6
Kaplan, D. (2018). Mechanisms and dynamical systems. In S. Glennan & P. Illari (Eds.), *The Routledge handbook of mechanisms and mechanical philosophy* (pp. 267–280). Routledge.
Kästner, L. (2021). Integration and the Mechanistic triad producing underlying and maintaining mechanistic explanations. In F. Calzavarini & M. Viola (Eds.), *Neural mechanisms studies in brain and mind* (Vol. 17). Springer.
Kersten, L. (2016). Commentary: The alleged coupling-constitution fallacy and the mature sciences. *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2016.02033
Kirchhoff, M. D. (2017). From mutual manipulation to cognitive extension: Challenges and implications. *Phenomenology and the Cognitive Sciences, 16*(5), 863–878. https://doi.org/10.1007/s11097-016-9483-x
Krickel, B. (2018). Saving the mutual manipulability account of constitutive relevance. *Studies in History and Philosophy of Science, 68*, 58–67. https://doi.org/10.1016/j.shpsa.2018.01.003
Kronfeldner, M. (2015). Reconstituting phenomena. In U. Mäki, I. Votsis, S. Ruphy, & G. Schurz (Eds.), *Recent developments in the philosophy of science: EPSA13 Helsinki* (pp. 169–181). Cham: Springer.
Kronz, F. M., & Tiehen, J. T. (2002). Emergence and quantum mechanics. *Philosophy of Science, 69*(2), 324–347. https://doi.org/10.1086/341056
Lee, J. (2021). Rise of the swamp creatures: Reflections on a mechanistic approach to content. *Philosophical Psychology*. https://doi.org/10.1080/09515089.2021.1918658
Lee, J., & Millar, B. (2022). Mechanisms of skilful interaction: Sensorimotor enactivism & mechanistic explanation [Unpublished manuscript].
Leuridan, B. (2012). Three problems for the mutual manipulability account of constitutive relevance in mechanisms. *The British Journal for the Philosophy of Science, 63*(2), 399–427. https://doi.org/10.1093/bjps/axr036
Levy, A., & Bechtel, W. (2020). Beyond machine-like mechanisms. In S. Holm & M. Serban (Eds.), *Philosophical perspectives on the engineering approach in biology* (pp. 99–122). Routledge.
Lyre, H. (2018). Structures, dynamics and mechanisms in neuroscience: An integrative account. *Synthese, 195*(12), 5141–5158. https://doi.org/10.1007/s11229-017-1616-4

Machamer, P., Darden, L., & Craver, C. F. (2000). Thinking about mechanisms. *Philosophy of Science, 67*(1), 1–25.
Maturana H. R. (1970). Biology of Cognition. Biological Computer Laboratory Research Report 9.0. University of Illinois, Urbana.
Maturana, H., & Varela, F. (1980). *Autopoiesis and Cognition*. Reidel.
McClamrock, R. (1991). Methodological individualism considered as a constitutive principle of scientific inquiry. *Philosophical Psychology, 4*(3), 343–354. https://doi.org/10.1080/09515089108573035
McGann, M., De Jaegher, H., & Di Paolo, E. (2013). Enaction and psychology. *Review of General Psychology, 17*(2), 203–209. https://doi.org/10.1037/a0032935
Miłkowski, M. (2016). Explanatory completeness and idealization in large brain simulations: A mechanistic perspective. *Synthese, 193*(5), 1457–1478. https://doi.org/10.1007/s11229-015-0731-3
Miłkowski, M., Clowes, R. W., Rucińska, Z., Przegalińska, A., Zawidzki, T., Gies, A., McGann, M., Afeltowicz, Ł, Wachowski, W., Stjernberg, F., Loughlin, V., & Hohol, M. (2018). From wide cognition to mechanisms: A silent revolution. *Frontiers in Psychology, 9*, 2393. https://doi.org/10.3389/fpsyg.2018.02393
Mill, J. S. (1843). *A System of Logic*. John W. Parker.
Okasha, S. (2002). *Philosophy of science: A very short introduction* (Vol. 67). Oxford Paperbacks.
Oppenheim, P., & Putnam, H., et al. (1958). The unity of science as a working hypothesis. In H. Feigl (Ed.), *Minnesota studies in the philosophy of science.* (Vol. 2). Minnesota University Press.
O'Regan, J. K., & Noë, A. (2001). A sensorimotor account of vision and visual conscsiousness. *Behavioral and Brain Sciences, 24*(5), 939–973. https://doi.org/10.1017/S0140525X01000115
Paoletti, M. P., & Orilia, F. (2017). Downward causation: An opinionated introduction. In M. Paolini Paoletti & F. Orilia (Eds.), *Philosophical and scientific perspectives on downward causation* (pp. 1–21). Routledge.
Piccinini, G. (2008). Computation without representation. *Philosophical Studies, 137*(2), 205–241. https://doi.org/10.1007/s11098-005-5385-4
Piccinini, G. (2020). *Neurocognitive mechanisms: Explaining biological cognition*. Oxford University Press.
Piccinini, G. (2022). Situated neural representations: Solving the problems of content. *Frontiers in Neurorobotics*. https://doi.org/10.3389/fnbot.2022.846979
Piccinini, G., & Craver, C. (2011). Integrating psychology and neuroscience: Functional analyses as mechanism sketches. *Synthese, 183*(3), 283–311. https://doi.org/10.1007/s11229-011-9898-4
Povich, M., & Craver, C. F. (2018). Mechanistic levels, reduction, and emergence 1. *The Routledge handbook of mechanisms and mechanical philosophy* (pp. 185–197). Routledge.
Ross, D., & Ladyman, J. (2010). The alleged coupling-constitution fallacy and the mature sciences. In R. Menary (Ed.), *The Extended Mind* (pp. 155–166). MIT Press.
Ruiz-Mirazo, K., & Moreno, A. (2004). Basic autonomy as a fundamental step in the synthesis of life. *Artificial Life, 10*(3), 235–259. https://doi.org/10.1162/1064546041255584
Salmon, W. C. (1984). *Scientific explanation and the causal structure of the world*. Princeton University Press.
Sheredos, B. (2021). Merleau-Ponty's implicit critique of the new mechanists. *Synthese, 198*(9), 2297–2321. https://doi.org/10.1007/s11229-018-02006-7
Silberstein, M., & Chemero, A. (2013). Constraints on localization and decomposition as explanatory strategies in the biological sciences. *Philosophy of Science, 80*(5), 958–970. https://doi.org/10.1086/674533
Smart, P. R. (2022). Toward a mechanistic account of extended cognition. *Philosophical Psychology*. https://doi.org/10.1080/09515089.2021.2023123
Sprevak, M. (2010). Computation, individuation, and the received view on representation. *Studies in History and Philosophy of Science Part A, 41*(3), 260–270. https://doi.org/10.1016/j.shpsa.2010.07.008
Stepp, N., Chemero, A., & Turvey, M. T. (2011). Philosophy for the rest of cognitive science. *Topics in Cognitive Science, 3*(2), 425–437. https://doi.org/10.1111/j.1756-8765.2011.01143.x
Stewart, J. R., Gapenne, O., & Di Paolo, E. A. (Eds.). (2010). *Enaction: Toward a new paradigm for cognitive science*. MIT Press.
Thelen, E., Schöner, G., Scheier, C., & Smith, L. B. (2001). The dynamics of embodiment: A field theory of infant perseverative reaching. *The Behavioral and Brain Sciences, 24*(1), 1–86. https://doi.org/10.1017/s0140525x01003910

Thompson, E. (2007). *Mind in life: Biology, phenomenology, and the sciences of mind*. Harvard University Press.
Thompson, E., & Stapleton, M. (2009). Making sense of sense-making: Reflections on enactive and extended mind theories. *Topoi, 28*(1), 23–30. https://doi.org/10.1007/s11245-008-9043-2
Van Gelder, T. (1995). What might cognition be, if not computation? *The Journal of Philosophy, 92*(7), 345–381. https://doi.org/10.2307/2941061
Varela, F. J., Thompson, E., & Rosch, E. (1991/2017). *The Embodied Mind: Cognitive science and human experience*. MA: MIT Press.
Vernazzani, A. (2014). Sensorimotor laws, mechanisms, and representations. Proceedings of the Annual Meeting of the Cognitive Science Society, 36. Retrieved from https://escholarship.org/uc/item/82v8d2dt
Vernazzani, A. (2019). The structure of sensorimotor explanation. *Synthese, 196*, 4527–5455. https://doi.org/10.1007/s11229-017-1664-9
Villalobos, M. (2013). Enactive cognitive science: Revisionism or revolution? *Adaptive Behavior, 21*(3), 159–167.
Villalobos, M., & Dewhurst, J. (2017). Why post-cognitivism does not (necessarily) entail anti-computationalism. *Adaptive Behavior, 25*(3), 117.
Villalobos, M., & Palacios, S. (2021). Autopoietic theory, enactivism, and their incommensurable marks of the cognitive. *Synthese, 198*(1), 71–87. https://doi.org/10.1007/s11229-019-02376-6
Villalobos, M., & Silverman, D. (2018). Extended functionalism, radical enactivism, and the autopoietic theory of cognition: Prospects for a full revolution in cognitive science. *Phenomenology and the Cognitive Sciences, 17*(4), 719–739. https://doi.org/10.1007/s11097-017-9542-y
Villalobos, M., & Ward, D. (2015). Living systems: Autonomy, autopoiesis and enaction. *Philosophy & Technology, 28*(2), 225–239. https://doi.org/10.1007/s13347-014-0154-y
Villalobos, M., & Ward, D. (2016). Lived experience and cognitive science: Reappraising enactivism's Jonasian turn. *Constructivist Foundations, 11*(2), 204–212.
Ward, D., Silverman, D., & Villalobos, M. (2017). Introduction: The varieties of enactivism. *Topoi, 36*(3), 365–375. https://doi.org/10.1007/s11245-017-9484-6
Wimsatt, W. C. (1994). The Ontology of complex systems: Levels of organization, perspectives, and causal thickets. *Canadian Journal of Philosophy Supplementary, 20*, 207–274. https://doi.org/10.1080/00455091.1994.10717400
Winning, J., & Bechtel, W. (2019). Being emergence VS. Pattern emergence: Complexity, control and goal-directedness in biological systems. In S. Gibb, R. F. Hendry, & T. Lancaster (Eds.), *The Routledge handbook of emergence* (pp. 134–144). Routledge.
Ylikoski, P. (2012). Micro, macro, and mechanisms. In H. Kincaid (Ed.), *The Oxford Handbook of Philosophy of Social Science* (pp. 21–45). Oxford University Press.
Zahavi, D. (2003). *Husserl's Phenomenology*. Stanford University Press.
Zednik, C. (2011). The nature of dynamical explanation. *Philosophy of Science, 78*(2), 238–263. https://doi.org/10.1086/659221

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.