THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

The ups and downs of mechanism realism

The Ups and Downs of Mechanism Realism: Functions, Levels, and Crosscutting Hierarchies

Joe Dewhurst[1] · Alistair. M. C. Isaac[2]

Abstract

Mechanism realists assert the existence of mechanisms as objective structures in the world, but their exact metaphysical commitments are unclear. We introduce *Local Hierarchy Realism* (LHR) as a substantive and plausible form of mechanism realism. The limits of LHR reveal a deep tension between two aspects of mechanists' explanatory strategy. *Functional decomposition* identifies locally relevant entities and activities, while these same entities and activities are also *embedded* in a nested hierarchy of levels. In principle, a functional decomposition may identify entities engaging in causal interactions that crosscut the hierarchical structure of composition relations, violating the mechanist's injunction against interlevel causation. We argue that this possibility is realized in the example of ephaptic coupling, a subsidiary process of neural computation that crosscuts the hierarchy derived from synaptic transmission. These considerations undermine the plausibility of LHR as a general view, yet LHR has the advantages that (i) its metaphysical implications are precisely stateable; (ii) the structure it identifies is not reducible to mere aggregate causation; and (iii) it clearly satisfies intuitive and informal definitions of mechanism. We conclude by assessing the prospects for a form of mechanism realism weaker than LHR that nevertheless satisfies all three of these requirements.

1 Introduction

Mechanistic explanation has emerged as a popular account of explanation across much of philosophy of science, especially for cognitive neuroscience, biology, and chemistry. This account suggests that to explain a phenomenon is to discover a mechanism that produces it. Crucially, for many mechanists, mechanisms differ from models and other idealized, counterfactual, or instrumental constructs of science, in

✉ Joe Dewhurst
    joseph.e.dewhurst@gmail.com

1   LMU Munich, Munich, Germany

2   University of Edinburgh, Edinburgh, UK

that mechanisms are actual parts of the world. Models, diagrams, simulations, or other descriptions may represent a mechanism, but the mechanism itself is a structure in the world, independent of our aims and interests, and it is this real structure that plays an explanatory role (either constitutively, for ontic accounts of explanation, or by reference, for epistemic accounts). Call this general position *mechanism realism*, i.e. realism about those features of a physical system referred to by a successful mechanistic explanation.

The exact metaphysical commitments of mechanism realism are unclear. Most generic definitions of mechanism commit the realist to (i) entities, performing (ii) activities, in a way that is (iii) organized (see e.g. Machamer et al. 2000; Glennan 2002; Bechtel and Abrahamsen 2005). While entities are a relatively unproblematic item in the metaphysician's toolkit, activities and organization are less standard, and they suggest two distinct routes for fleshing out a metaphysics of mechanisms. One route takes activities to be a new ontological category, distinct from causal interactions understood in a generic sense. This move is suggested by Machamer et al. (2000), and has been influential in distinguishing their view from that of Glennan (1996). Nevertheless, it's a strategy that faces challenges, as even proponents find it difficult to determine necessary and sufficient conditions for an activity (Machamer 2004; cf. Krickel 2018), and detractors argue that mechanists must grant both properties and counterfactuals, rendering activities supererogatory (Psillos 2004).

Here, guided by our own preference for a desert ontology, we explore an alternative route toward a realist metaphysics of mechanisms. We assume activities are mere causal interactions, and ask whether there is a distinctive *structural* constraint placed on entities and activities by the mandate that mechanisms be 'organized'. We take it as a success condition for this project that a distinctively mechanistic mode of organization identify some arrangements of entities and their interactions as mechanisms, and rule out others as not. Thus we reject a view on which attributing 'mechanistic' structure to a system amounts to no more than attributing causal structure (e.g. Wilhelm 2019). A central tenet of the mechanist movement is that mechanistic explanation is distinct from, and more metaphysically ambitious than, explanations that merely aggregate causes (see e.g. Craver and Bechtel 2007; Craver 2015; Darden 2002; Raerinne 2011). If mechanistic explanation is distinct from causal explanation, then realists about mechanisms should hold out for a metaphysics that does not merely reduce mechanisms to a collection of causes.

The remainder of this paper elucidates and critiques one *prima facie* plausible way to characterize the distinctive structure of mechanisms, which we call *Local Hierarchy Realism* (LHR). Although ultimately we abandon LHR as untenable, we think it is constructive to consider it seriously for several reasons. First, LHR is intuitive and easy to visualize, and plausibly informs much of the construction of mechanistic explanation in the sciences—if we reject LHR, then we should also be wary of intuitive discussions of mechanism structure, or of putting any ontological weight on diagrams and other easy to digest representations of mechanisms.[1] Second, LHR

---

[1] Our claim is not that mechanistic explanations in general satisfy LHR, but rather that some intuitive understanding of LHR drives the search for mechanistic explanation. This is an empirical claim that goes beyond the scope of this paper, but to see why it might be plausible, notice that diagrams and intuitive visualizations satisfying LHR play a substantive role in mechanism discovery (Sheredos et al. 2013), and

treats hierarchical structure as highly localized and pluralistic, and thereby serves as a bridge between discussions of hierarchy in philosophy of science and in metaphysics. Where philosophers of science have rejected the thoroughgoing hierarchy realism of Putnam and Oppenheim (1958), metaphysicians have attempted to save some tempered version of it, as debates about fundamentality, grounding, and emergence all seem to presuppose a stratification of nature into levels (Kim 2002). The challenge we raise for LHR, however, concerns the possibility of crosscutting levels, and thus also challenges the tempered versions of levels advanced so far amongst metaphysicians (e.g. List 2019). Third, formalizing LHR suggests specific ways forward for a structural account of the metaphysics of mechanisms, and indeed of levels-talk more broadly.

The final, and perhaps most important, reason to examine LHR is that its limitations reveal an unrecognized tension between two aspects of mechanistic explanation that serve to distinguish mechanisms from mere aggregate causes. On the one hand, mechanisms are typically understood as mechanisms *for* the production of some phenomenon (Machamer et al. 2000: 3; Glennan 2002: S344; Bechtel and Abrahamsen 2005: 423; Illari and Williamson 2012: 120), and attributing the *function* of producing this phenomenon to a mechanism identifies and individuates it (Garson 2013; Shagrir and Bechtel 2017). More generally, attributing functions to a mechanism as a whole and subsidiary functions to its parts serves to identify a substantive organizational structure over and above mere causal aggregation. On the other hand, the construction of an effective mechanistic explanation is typically taken to involve both 'looking down' at the internal structure of the components of a mechanism and 'looking up' at the broader pattern of activities in which a mechanism as a whole participates (Bechtel 2009). The resulting picture is one of a hierarchy of multiply embedded mechanisms, with component entities at any one level themselves treated as mechanisms at a finer grain, and decomposed into lower-level entities and activities. Likewise, any mechanism may itself be treated as a unitary component, or an entity in a higher-level, more coarse-grained mechanism (Craver 2001, 2015; Glennan 2005). The stratification of mechanism entities and activities into a local hierarchy of levels also suggests a substantive organization that may not be reduced to mere aggregate causes (Craver 2015; Craver and Bechtel 2007).[2]

We demonstrate that the decomposition of a mechanism into entities and activities through function attribution, and the subsequent arrangement of mechanistic

---

Footnote 1 (continued)

also that, anecdotally, scientists studying mechanisms often have only a hazy view of what a mechanism amounts to, and for this very reason have found it productive to collaborate with philosophers, for example, molecular biologist John Moult's collaboration with philosopher Lindley Darden on the MecCog mechanism representation system (Kundu et al. 2020).

[2] In what follows, we draw on influential comments from prominent figures in the mechanisms literature, such as Craver, Bechtel, and Glennan. We do not intend to attribute LHR, nor indeed any specific metaphysical view, to any of these authors individually. Rather, we think the views they separately articulate represent commitments that large segments of the mechanism community endorse, or at least take seriously. Consequently, it is worthwhile to investigate the metaphysical implications of these claims taken as a whole.

components into a hierarchy of levels, place inconsistent demands on mechanism structure. This follows as an a priori possibility from the observation that the different functional roles contributing to a process for generating some phenomenon may induce 'crosscutting' hierarchies of causal relations (Sect. 3). We add empirical weight to this a priori argument through the example of ephaptic coupling, a functional mechanism in neuroscience that contributes to the phenomenon of neural firing, and appears to undermine realism about the hierarchical mechanistic structure of neuron–neuron interactions performing neural computation (Sect. 4). Finally, we articulate the formal conditions LHR places on the interaction between composition and causality (Sect. 5). This formulation of the metaphysical commitments of LHR suggests a way forward for precisely articulating weaker forms of mechanism realism that nevertheless make substantive claims about the organization of entities and activities in the world.

We conclude by posing a challenge for the structural approach to the metaphysics of mechanisms. On the one hand, LHR has strong intuitive appeal and is easy to visualize, but these appealing features may only be preserved by mechanism realists if they can find some resolution to the apparent tension between functional decomposition and hierarchical mechanism construction. On the other hand, abandoning LHR leaves a large and relatively unexplored space of metaphysical possibilities richer than mere aggregation of causes. Nevertheless, these possibilities are difficult to grasp intuitively, and do not obviously conform to the qualitative understanding of mechanism derived from scientific explanations. Consequently, it seems the burden falls on the would-be non-LHR mechanism realist to defend some weaker account of the ontological implications of mechanistic explanation as worthy of the name 'mechanism.'

2 Function Attribution and Local Hierarchy Realism

In this section, we motivate Local Hierarchy Realism by looking at the process by which mechanistic explanations are constructed, and clarify the sense in which plausible mechanism hierarchies should be understood as purely local in scope.

Mechanistic explanation typically begins by identifying a target phenomenon (cf. Shagrir and Bechtel 2017), and positing a mechanism with the *function* of producing that phenomenon (Garson 2013). The components of a mechanism are often also mechanisms in their own right, performing subsidiary functions that contribute to the overall function of the top-level mechanism, and in general attributing functions to both a mechanism and its parts appears necessary for recovering a non-trivial notion of mechanistic organization. Nevertheless, mechanism realists typically do not wish to include functions as part of their primitive ontology, but rather explain away, reduce, or derive functions from evolutionary history (Millikan 1989; Neander 1991), selective pressures (Garson 2017), or "objective goals" (Maley and Piccinini 2017). On the most deflated view of function attribution, it serves merely as a heuristic, capturing the perspective of the researcher on the target phenomenon (Craver 2013; Cummins 1975; Darden 2008; Glennan 2017; Hardcastle 1999; Kästner 2018).
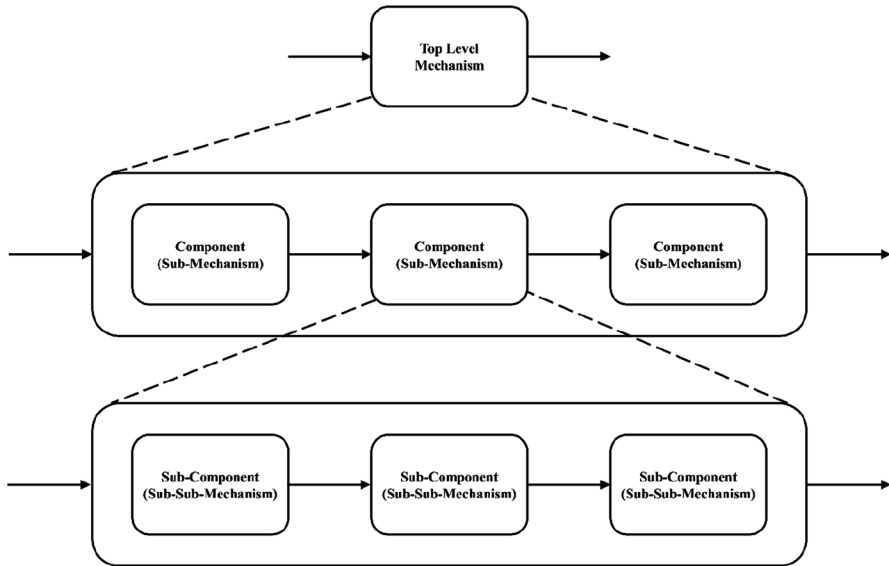
Fig. 1 A schematic version of the hierarchical structure of mechanisms. Black arrows indicate causal interactions, whereas dashed lines indicate the composition relation between levels. As typically understood, there can be no causal interaction between levels, and a level is defined in terms of the causal interactions between its parts (components)

Function attribution and functional analysis serve to isolate some set of entities and identify the characteristic activities they engage in to produce a target phenomenon. However, isolating these entities and identifying their characteristic activities is not yet enough to show they constitute more than mere aggregate causes. On the contrary, the deflated view of functions explicitly asserts that the interactions between mechanistic entities are merely causal interactions. Yet the process of functional analysis does suggest a substantive metaphysical constraint, once we examine how it relates mechanisms to one another. Functional analysis iteratively decomposes a complex system into a hierarchy of subsystems arranged into levels of components at increasingly finer grain, where each level of this hierarchy may serve as a "mechanism sketch" (Piccinini and Craver 2011; Shagrir and Bechtel 2017). The upshot is that mechanisms are arranged into hierarchies, such that the constituent entities (components) of any mechanism may themselves be treated as mechanisms at a finer grain, and decomposed into lower-level entities and activities. Moving up the hierarchy, any mechanism may itself be treated as a unitary component, or an entity in an even higher-level, more coarse-grained mechanism (Bechtel 2009). The picture is one of arbitrarily many nested levels of entities, each engaging in characteristic activities at some fixed level of the hierarchy (Fig. 1; Craver 2001, 2015; Glennan 2005).

To see how this works, consider the oft-discussed example of the heart. The rhythmic contraction of the heart serves as a target phenomenon: how might we explain this? A mechanistic explanation must start from a (perhaps tentative) attribution of function. For instance, William Harvey's insight that the heart could be

understood as functioning analogous to a pump identifies component entities, such as the right ventricle and aortic valve, and assigns them subsidiary functions, such as expelling blood and preventing blood flow from reversing. One might then take one of these entities, say the right ventricle, and provide a mechanistic explanation of *its* characteristic behavior. Here, the subsidiary function of expelling blood assigned to the ventricle at a higher level, becomes the primary function that then serves to define subsidiary functions to its component entities, such as the contraction function of muscle cells and the joining function of the intercalated disks. Conversely, one might step back from the heart as a whole, and consider its pumping as a subsidiary function within the broader mechanism of the circulatory system. In general, analyses of mechanistic function cascade through levels of a hierarchy, as the attribution of a functional role to an entity at one level of the hierarchy will both determine which entities and activities are identified as components of that entity's mechanism at the next level down, and constrain the kinds of broader functional role that entity can perform at the next level up (cf. Cummins 1975).

Call realism about this hierarchical arrangement of mechanisms into levels 'Local Hierarchy Realism', or LHR. LHR is the view that, once we have specified a target phenomenon, there is some fact of the matter about the hierarchical structure of the mechanism that explains it, i.e. how it ought to be decomposed and which components and subcomponents will appear at which level of the hierarchy. LHR is a metaphysically substantive view that does not reduce to a mere aggregation of causes. Our claim is not that any particular mechanist has explicitly defended LHR, but rather that it follows naturally from the generic account of mechanistic explanation as articulated by philosophers of mechanism, and that it has intuitive appeal as an easy to visualize metaphysical picture, a straightforward reification of common mechanism diagrams like Fig. 1. To better flesh out LHR, let's look at two possible points of confusion.

First, it is crucial to stress that LHR is a thesis about *both* mereology and causation. At first blush, it may seem that the talk of entities decomposing into mechanisms at lower levels implies only a mereological hierarchy. However, the 'activities' within each level of a mechanism should be understood as highlighting the relevant aspects of its internal causal structure (i.e., for producing the target phenomenon), and because mechanists typically take causation to be strictly intra-level, this structure is also constrained by LHR. This means that a mechanistic analysis of causal structure is also going to be an analysis of hierarchical structure, and vice versa (we make this claim more precise and explore its metaphysical implications in Sect. 5). The strict commitment to accepting only intra-level causes is confirmed in the complete rejection of inter-level causation by many mechanists (Baumgartner and Casini 2017; Craver and Bechtel 2007; Kaiser and Krickel 2017; Romero 2015). Most notably, Craver and Bechtel (2007) argue that any putative cross-level causation is always reducible "without remainder" to a combination of composition relations across levels and intra-level causation. For instance, the saturation of photopigment in cone cells does not cause the eye to transduce light into neural signals; rather, photopigment is a constitutive, lower-level component of the mechanism of the eye, which as a whole, at a higher level, participates in an intra-level causal interaction between light and electrical activity in the brain (Craver and Bechtel 2007: 555). We

call this the Craver–Bechtel (C–B) principle. The C–B principle provides LHR with a principled approach to subsuming putative counterexamples into a single mechanistic hierarchy, by explaining away apparent inter-level causes in terms of compositional relations and intra-level causes.

Second, in order to be plausible, LHR should not imply *universal* hierarchy realism (UHR). UHR holds that the entirety of nature is neatly organized into a hierarchy of levels of scale and interaction, a view most famously endorsed by Putnam and Oppenheim (1958). Although once widely held, this view has been largely discredited by detailed scientific examples, especially from the biological sciences (e.g. Potochnik 2017). Any plausible form of realism about levels should accept that mechanisms are a purely local affair, and thus that their hierarchical structure is also purely local. In fact, many mechanists are strongly opposed to making any claims at all about global hierarchies, or about how the levels of one mechanistic hierarchy might relate to those of another mechanism (e.g. Povich and Craver 2017; Eronen 2015; however argues such stringent locality does not capture how levels are actually described in the biological sciences). There are two distinct ways of interpreting the claim that mechanism hierarchies are local: they could be functionally local or they could be spatiotemporally local. Spatiotemporal locality is satisfied if a hierarchical arrangement of entities is tightly circumscribed by a well-defined spatiotemporal boundary, such as the hierarchy of organelles, proteins, and amino acids within a cell (cf. Craver 2007: 191). Functional locality is satisfied when a hierarchy is restricted to levels emerging through decomposition with respect to a tightly circumscribed functional attribution; for instance, sodium ions in the brain may be organized in levels with different sets of entities depending on whether their functional role in metabolism or in neural computation is under investigation. Either type of locality would be enough to prevent LHR from collapsing into UHR, but we think the most plausible form of LHR should take the most stringent route, and attribute reality only to the levels of mechanisms that are both functionally and spatiotemporally localized.

3 The Crosscutting Threat to Local Hierarchy Realism

LHR derives its metaphysical commitments from an explanatory practice that iteratively embeds mechanism levels through cascading function attributions. Yet this practice only contingently produces a consistent hierarchy, and in principle allows subsidiary function attributions to produce levels that crosscut one another. This kind of crosscutting would be inconsistent with the central claim of LHR, i.e. that each mechanism possesses a *unique* hierarchical structure about which we ought to be realists.

Our core argument is as follows. Functional decomposition specifies intra-level causal structure. Once intra-level causal relations are fixed, inter-level composition relations are constrained to induce a hierarchy of levels. However, a single phenomenon may in principle induce multiple functional decompositions, and competing function attributions may fix different causal relations as intra-level, which in turn induce their own hierarchies of composition relations. Since hierarchical structure

is derived from intra-level causal structure (and not vice versa), and since intra-level causal structure is fixed by function attribution, functionally local mechanistic hierarchies are vulnerable to the possibility that causal relations picked out by subsidiary function attributions crosscut one another. In principle these functionally local crosscutting hierarchies may be spatiotemporally local as well, thereby threatening LHR. The remainder of this section elaborates some contentious steps in this argument in more detail.

Function attribution specifies intra-level causation. Suppose we have specified a target phenomenon and need to determine the structure of the mechanism that produces it through functional decomposition. Craver identifies three kinds of function attribution in mechanistic model building: etiological, input–output (I–O), and contextual. It is the second of these that is critical for determining intra-level causation, namely "function [as] a mapping from inputs to outputs in conformity with a rule" (2013: 149). To attribute an I–O function to a putative mechanism component is to identity some causal influences ('inputs') and some characteristic activities ('outputs') as constitutive of that component's role in producing the explanandum phenomenon. This I–O attribution assigns these putative causal relations a special status in the context of model construction, because the C–B principle admits only intra-level causal relations within a single mechanistic system. Since the entities that input to a component, the activities that component outputs, and the entities impacted by those activities all interact causally, they are thereby constrained to fall on the same level of the mechanism, as ontological (and compositional) equals.[3]

Intra-level causal relations induce mechanistic hierarchy. Once a single level of a mechanism is fixed, the complementary processes of 'looking up' and 'looking down' elaborate it into a hierarchy, in a manner heavily constrained by the causal structure functionally attributed to the reference level. Looking down, each component of the reference level that performs an I–O function may itself be broken down into subcomponents performing subsidiary functions. Crucially, the spatial and functional scope of these subcomponents within the mechanism is constrained by the initial analysis:

> Isolated descriptions of an X's ɸ-ing specify the activity for which a lower-level mechanism will be sought and so fix the active, spatial, and temporal boundaries of that mechanism. (Craver 2001: 65; cf. 2013: 151)

Likewise, looking up, one may treat a pattern of causally interacting entities at one level as constituting a single entity (component) at a higher level. This higher-level entity itself performs an I–O function, analyzable in terms of intra-level causal relations at that higher level (Bechtel 2009).

---

[3] This point should also clarify our use of the C-B principle: we take it to be a *symmetrical* constraint that holds between composition and causal relations. Conversely, one might take it to be an asymmetrical constraint, for instance, constraining causal relations on the assumption of composition relations (but not vice versa). This reading has been suggested to us by a reviewer; however, our own view is that, provided one assumes that mechanisms must be functionally local, the arguments of Craver and Bechtel (2007) are strong enough to motivate the symmetrical principle as well.

It is critical for what follows to recognize that, while composition relations respect a mechanistic hierarchy, the stratification of a mechanistic explanation into levels is driven in the first instance by function attribution. Here we disagree with Craver, who suggests in places that.

> Levels of mechanisms are defined fundamentally by . . . the componency relationship between things at higher and lower levels. (Craver 2015: 19)

In contrast, we take compositional relations at lower and higher levels to be constrained by the entities at some initial level. Since these initial ontological boundaries are themselves determined by a prior functional analysis, hierarchical structure is ultimately determined by function attribution. Furthermore, function attribution, in the form of subsidiary functions derived from the decomposition of the higher-level mechanism function (or the upwards embedding of a lower-level mechanism function), continues to play the driving role in level membership, since it picks out relevant causal and ontological structure at subsequent levels, ensuring that hierarchical stratification is functionally local (cf. Cummins 1975).

Since this is a contentious point, it's worth reflecting in more detail on why functional analysis is prior to mereological analysis: couldn't one first parse the putative mechanism into relevant parts, and only after attribute functions to those parts? If that were the case, then compositional structure would be independent of functional analysis. There is a strong intuition that beginning from mereological analysis should be possible. Consider a paradigmatic mechanism, e.g. a grandfather clock: aren't the gears, chains, weights, springs, face, and cabinet all easily identifiable as hierarchically arranged components, even if we have no idea whatsoever how a clock works, or what it is for? But this intuition it misguided, for it ignores the substantive role played by judgments of similarity and relevance in any such decomposition. Why is the suggested breakdown different from one that decomposes the clock into left and right halves, or metal components and wood components, or small parts and large parts? In fact there are too many 'natural' ways to decompose the clock into a hierarchy of parts, each assuming a different answer to the question of what features of its physical composition are relevant for determining part boundaries and composition relations—mereology is wholly unconstrained without some principle of relevance by which to distinguish those properties that contribute to part individuation from those that do not. Functional analysis provides this principle of relevance, at least so far as mechanistic explanation is concerned. Attributing the function of doorstop, or *objet d'art*, or timekeeping device to the clock will distinguish different properties as relevant, thereby initiating different mereological decompositions.

Crosscutting hierarchies undermine LHR. So, functional analysis determines the stratification of a putative set of embedded mechanism components into levels. Since hierarchy realism is local, in the sense that it only applies to hierarchies that are both spatiotemporally and functionally localized, one might think that there is no danger of inconsistencies across the mechanism levels that are strung together by a cascading process of functional decomposition. In particular, although at each stage of this process new function attributions are made, we might expect the fragments of mechanism hierarchy they induce to always be embedded or otherwise combined with other local fragments to form a single, more elaborate localized hierarchy, one

that we can maintain a straightforwardly realist attitude towards. This realist interpretation is threatened by the possibility that levels may *crosscut* one another, i.e. that a single functional analysis might induce two (or more) distinct decompositions, such that token entities occur at the same level relative to one decomposition and at different levels relative to another. If a single functional analysis could in principle produce crosscutting hierarchies within a spatiotemporally localized system, then one could not infer any unique hierarchical organization for this system, not even of a locally circumscribed sort. More generally, the hierarchical organization of any particular mechanistic model would not guarantee the uniqueness of that hierarchical organization in the world, thus threatening LHR.

Nevertheless, the process of functional decomposition does seem to allow for crosscutting hierarchies of the same functionally individuated and spatiotemporally circumscribed components. The key point is just that the assignment of an I–O function to a particular entity has rippling consequences, spreading out from that entity to identify the other members of its level of causal interaction. The functional decomposition of an entity at one level into its component parts then assigns these parts subsidiary functions, but nothing guarantees that the subsidiary function of an entity at one level might not ripple out and engage in direct causal interactions with entities already localized by its parent function at some different level. Causal interactions across levels such as this pose a challenge to LHR, as they induce new levels that 'crosscut' the existing hierarchy. The primary response of the hierarchy realist against this possibility is the application of the C–B principle: any putative interlevel cause may be decomposed into a chain of intra-level causes and componency relations. But the C–B principle cuts both ways—given two crosscutting hierarchies, a defender of either may adopt the C–B principle to explain away its competitor's putative causes. Yet this power to translate causal structure between multiple hierarchies undermines realism about any particular hierarchical organization.

Here's an abstract description of the worry: suppose that a functional decomposition F assigns subsidiary I–O functions s and t to parts of a spatiotemporally localized system $\mathbf{X}$. Call the layers of causal interaction induced by s and t, S and T respectively. Now suppose that S and T crosscut each other, i.e. by the lights of $S(\mathbf{X})$, **a** causally interacts with **b** and **c** is a lower level component of **a**, while by the lights of $T(\mathbf{X})$, **b** causally interacts with **c**, which is itself still a component of the higher level entity **a** (Fig. 2). Taking inspiration from Craver and Bechtel, we can define the *apparent* causal relations in a mechanism $\mathscr{C}$ (•) as the transitive closure of any path of cause and composition relations (in either direction). Then it is easy to see that $\mathscr{C}(S(\mathbf{X})) = \mathscr{C}(T(\mathbf{X}))$, and consequently the two hierarchies are consistent as analyses of apparent causal structure. But the two hierarchies also crosscut one another, so the upshot is that I–O analysis by itself is insufficient to derive a unique mechanistic hierarchy, at most it can resolve apparent causal structure without specifying which hierarchy is 'real'.

Function attributions determine hierarchical structure, but in principle the process of functional decomposition may produce crosscutting causal hierarchies within a mechanistic explanation of a single phenomenon. The C–B principle saves this picture from inconsistency, by showing that apparent crosscutting causes may be made consistent. The cost of consistency, though, is abandoning the idea of a unique
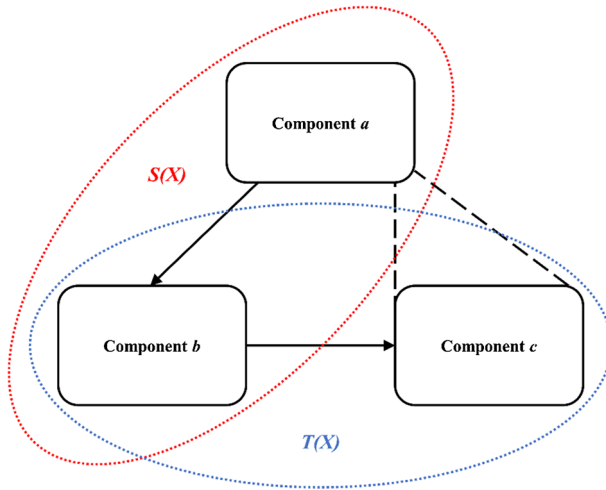
Fig. 2 Component **c** is a sub-component of component **a**. Components **a** and **b** causally interact for $S(\mathbf{X})$, and thus sit at the same mechanistic level, while **b** and **c** causally interact, and thus occupy the same level, for $T(\mathbf{X})$. Therefore the componential hierarchy of one function attribution, $S(\mathbf{X})$, cross-cuts that of another, $T(\mathbf{X})$

hierarchical decomposition of the system, even in a tightly circumscribed and local sense. Since LHR is the view that, at least locally, there is a unique hierarchical decomposition of any mechanism, this shows that LHR doesn't follow as a matter of course from realism about mechanistic explanations—i.e. brute realism about mechanisms as scientific explanations isn't enough to ensure a general metaphysical picture that is richer than mere aggregate causes.

This is an in-principle challenge to LHR, but should it worry a naturalistic metaphysician? Perhaps localized crosscutting hierarchies are logically possible but nomologically impossible? Or perhaps the actual practice of scientists when functionally decomposing complex phenomena prohibits crosscutting hierarchies? In the next section we present a concrete example that demonstrates crosscutting is a genuine empirical possibility, and thus a serious threat to local hierarchy realism.

4 Crosscutting Hierarchies in Practice

In this section we examine one apparent example of crosscutting mechanism hierarchies derived through analysis of a single overarching phenomenon. While we suspect that crosscutting hierarchies may be common in certain explanatory domains (especially those of the biological and social sciences), all that is needed to block the naturalistic defense of LHR is a single counterexample. The challenge for any putative counterexample is to ensure it satisfies the locality constraints on LHR, and we conclude the section by comparing our example to a *prima facie* similar case where locality is indeed violated, in order to tease out the sense in which our example poses a genuine challenge to LHR.
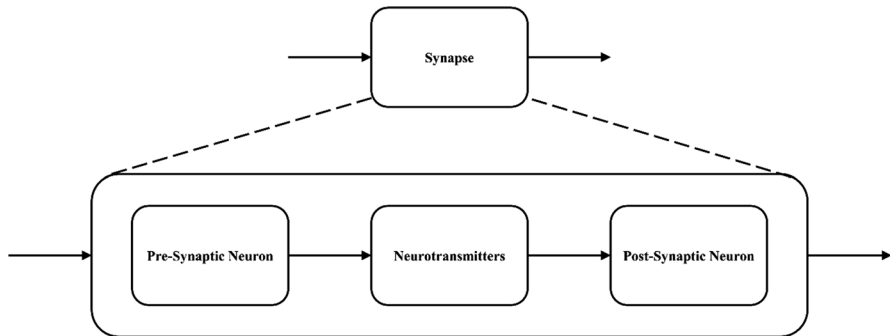
Fig. 3 A schematic decomposition of the local hierarchy responsible for synaptic transmission, where black arrows indicate causal interactions, and dashed lines indicate the constitution relation between levels of the hierarchy

The single target phenomenon at issue is neural computation, and in particular the inducement of a particular token neuron to fire as part of a neural computation. Neural computation is typically explained in terms of a mechanistic hierarchy, with neurons arrayed at a single level, engaging in a characteristic activity of inducing each other to fire through synaptic transmission (Boone and Piccinini 2016; Elber-Dorozko and Shagrir 2019; Piccinini and Bahar 2013). Looking down into individual neurons, the action potential explains the internal mechanism by which firing occurs; looking up, neurons are embedded in broad networks, as determined by the pattern of connection between axons and dendrites; families of neurons and entire brain regions may likewise be identified with higher organizational levels of neural computation (Kaiser 2007). However, the very same phenomenon, the firing of a particular neuron as part of the process of neural computation, may also be induced through a process distinct from synaptic transmission, namely ephaptic coupling. When we examine the mechanistic levels induced by synaptic transmission and ephaptic coupling, we discover they crosscut each other—token-identical entities inhabiting a single level in the first decomposition sit at different levels in the second. This example poses a challenge to LHR as it seems that spatiotemporally localized entities (i.e. two token-identical neurons) may interact in a manner derived from a single functional decomposition (namely, neural computational role) yet which does not result in a unique arrangement of mechanistic levels.

The standard view of neural computation identifies neural firing with the transmission of information in aid of a computation (e.g. Piccinini and Bahar 2013). This basic picture is neutral across debates about the exact way in which neural firing encodes information, namely whether individual spikes in neural activity, rates of spiking, sequences or trains of spikes, or spatial or probabilistic distributions over larger-scale patterns of firing should be identified as the vehicles of information. Crucial for any of these views is the mechanism by which one neuron causally influences (exciting or inhibiting) the firing of another. The most studied route for this interaction is chemical synaptic transmission (Fig. 3; cf. Shepherd 1994; Craver 2007: 135–8; Piccinini and Craver 2011). An action potential travels down the axon

of the pre-synaptic neuron, Ca^{2+} enters the end of the axon, inducing the pre-synaptic neuron to release neurotransmitters, which in turn cross the synaptic cleft to bind to receptors on the post-synaptic neuron's dendrite. Depending on the nature of the neurotransmitters, their effect on the post-synaptic cell will either inhibit its firing, or sum with other excitatory signals to induce it to fire itself.

Beginning from an analysis of neural computation as information processing, the subsidiary function in this example is transfer of information from one cell to another. Identifying this function isolates the pre-synaptic neuron, the post-synaptic neuron, and the molecular neurotransmitters released by the first and binding to the other as all interacting causally, and thus at the same level of a mechanistic hierarchy. The interaction between these components produces the overarching phenomenon of synaptic transmission at a higher level of this hierarchy.

Now consider another, less well-known process that also contributes to neural firing: ephaptic coupling. Ephaptic coupling occurs when an electromagnetic field affects the firing rate of an individual neuron—in a typical case, a field itself generated by the electromagnetic activity of another group of neurons (see e.g. Anastassiou et al. 2011). The possibility of ephaptic coupling was experimentally demonstrated in the 1940's (Arvanitaki 1942), and has since then been studied both in vitro and in vivo. A point of continued contention concerns whether ephaptic coupling plays any significant role in normal brain function. It is relatively uncontentious that it can play a role in pathological conditions—for instance, ephaptic coupling amplifies random activity, allowing disruptive neural "cross-talk" between demyelinated neurons in trigeminal neuralgia (Love and Coakham 2001). Nevertheless, there is also substantial evidence for a non-pathological role; for instance, it has been demonstrated in vivo that ephaptic coupling affects computationally significant features of neurons, such as spike timing and threshold of neural response (Anastassiou et al. 2011); furthermore, recent work has offered evidence that ephaptic coupling functions to regulate periodic neural activity (Chiang et al. 2019). Low frequency periodic activity in the hippocampus is believed to play a role in memory consolidation (Dickson 2010). Chiang et al. (2019) observed the propagation of low frequency periodic neural activity of this sort in hippocampal tissue that had been severed in order to ensure chemical synaptic transmission played no role. They then developed a computer model demonstrating how this propagation may be explained through ephaptic coupling induced by an endogenous electromagnetic field (Chiang et al. 2019: 261). Their study seems to be a clear case of mechanistic explanation, decomposing a system into components and their characteristic interactions, and yet it generates a causal hierarchy that crosscuts the standard synaptic transmission mechanism outlined above. If ephaptic coupling indeed plays a functional role in neural computation, then this crosscutting hierarchy poses a challenge to LHR.

In order to see why, we need to examine the mechanistic structure of ephaptic coupling. An endogenous electromagnetic (EM) field is generated by the collective behavior of a group of neurons, and is not reducible to the behavior of individual neurons, nor a mere aggregate of that behavior. Rather it is a holistic phenomenon, depending for its constitutive properties on the neurons' collective interaction. In other words, the EM field occurs at a higher level of the mechanistic hierarchy than the generating neurons themselves. However, the causal interaction at issue occurs
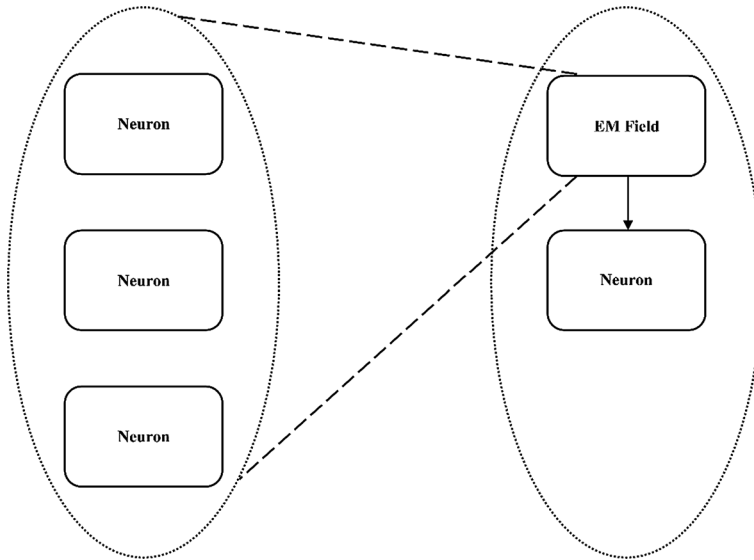
Fig. 4 A schematic decomposition of the mechanism hierarchy responsible for ephaptic coupling. As in the previous diagram, the black arrow indicates a causal interaction between an electromagnetic field and a neuron, while the dashed lines indicate that the field is constituted by the activity of some other set of neurons, which consequently sit at a lower level of the hierarchy. Each dotted oval designates a distinct level of the ephaptic coupling hierarchy, presented left-to-right rather than top-down for reasons that will soon become clear

between this EM field and a single, target neuron, the one entrained by it. In other words, because ephaptic coupling occurs between an EM field and a neuron, both that neuron and the field should appear on the same level of the corresponding mechanistic hierarchy (Fig. 4).

In principle, both of these mechanisms (synaptic transmission and ephaptic coupling) may be active at the same time, involving the same token neurons, satisfying spatiotemporal locality. Furthermore, they may both contribute to the computationally significant firing pattern of a single neuron, and may need to be invoked simultaneously to provide a full mechanistic explanation of that phenomenon, satisfying functional locality.[4] So, a single overarching phenomenon decomposes into

---

[4] Is this example only a problem for a structural account of mechanism realism; would mechanism realists who countenance activities as a distinct ontological category fare better? Not necessarily: Persson (2010) argues that activities-based accounts can't adequately model polygenic causation, cases where two distinct causes contribute to a single effect. Plausibly his arguments apply to a case like this, where both ephaptic coupling and synaptic transmission may contribute to a single neuron spike train. The case becomes even more problematic if we compare *electrical* synaptic transmission—whereby the electrical activity of a neuron induces firing across a synaptic gap, rather than the chemical release of neurotransmitters—and ephaptic coupling. In this case, it seems the very same activity of the pre-synaptic neuron (generating a local electrical signal) may causally affect the firing of the post-synaptic neuron through two distinct routes, both directly and by means of a higher-level EM field. We do not discuss this example in the main text as we are not aware of any empirical work on the interaction between these two processes.
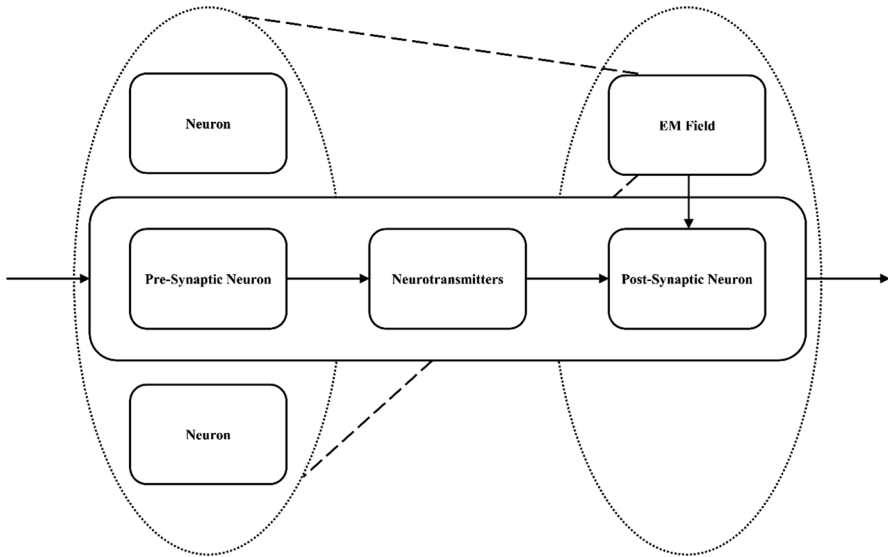
Fig. 5 The two mechanism hierarchies crosscutting one another. Here the pre-synaptic neuron is part of the group of neurons whose activity constitutes the electromagnetic field, which in turn exerts a causal influence on the post-synaptic neuron. Entities at a single level of the synaptic hierarchy appear at two different levels of the ephaptic hierarchy

subsidiary functions that induce two distinct mechanistic hierarchies that crosscut one another, with a set of components that are situated at one level in the first hierarchy appearing at different levels in the second. Relative to the functional decomposition of synaptic transmission, the pre- and post-synaptic neurons are situated at the same mechanistic level, as demonstrated by the fact that they interact causally with one another. Relative to the ephaptic coupling functional decomposition, however, these very same neurons might be situated at distinct levels, as the electromagnetic field acting on the post-synaptic neuron may be constituted by the firing of some other set of neurons that includes the pre-synaptic neuron (Fig. 5). In this case the pre-synaptic neuron is both involved in generating an electromagnetic field and producing neurotransmitters that contribute to the firing of the post-synaptic neuron, thereby simultaneously occupying two mechanistically distinct 'levels'. Thus, there is no 'objective' fact about whether these neurons are on the same or different levels, as their relative positions in the hierarchy change with the attribution of different subsidiary functions in the analysis, and so neural computation stands as a counter-example to LHR.

To tease out what is distinctive about this example, it will be useful to contrast it with a prima facie similar case that has been specifically disavowed by localist accounts of mechanistic levels.[5] Awake craniotomy is a form of brain surgery where

---

[5] We thank one of the referees for suggesting this example, which is adapted from Bechtel (2006).

the neurosurgeon operates on a patient under local anaesthetic, while also talking to them, or otherwise engaging their higher mental functions (a musician might be asked to sing or play their instrument during the procedure, for instance; Bonhomme et al. 2009; Scerrati 2020). The surgeon's motor regions control her movements, which themselves causally affect the patient's brain as, say, she removes a tumor from near Broca's area—i.e. the surgeon's brain and the patient's brain appear at very different levels of any putative organizational hierarchy. Yet, the surgeon may be talking to the patient at the very same time, to ensure that no damage is done to the patient's speech as the tumor is removed—i.e. relevant regions of the surgeon's brain and of the patient's, such as their respective Broca's areas, seem at the same level as they contribute symmetrically to explaining this conversation. Nevertheless, a localist account of mechanistic levels rejects the appearance of contradiction here, and asserts that this example poses no principled challenge to a sufficiently local understanding of mechanism hierarchy realism (cf. Bechtel 2006: Ch. 2). If awake craniotomy and neural computation are closely analogous, it seems the example of ephaptic coupling cannot charitably be taken to challenge LHR.

In our view, the critical difference between the two cases is that awake craniotomy does not satisfy the functional locality constraint, while ephaptic coupling, understood through its functional role in neural computation, does. To see this, consider what in the awake craniotomy case may count as the single phenomenon to be explained (analogous to the firing of a single neuron), and the single overarching functional analysis (analogous to computational role) that results in crosscutting levels. Plausibly, the phenomenon to be explained is the speech uttered by the patient during surgery; yet, it is then difficult to see how the modulations in this speech induced by conversation with the surgeon and those induced by the interventions of her scalpel may be subsumed under the same functional analysis. Aspects of the patient's speech induced conversationally are typically meant as checks on normal functioning—patients are asked to name pictured objects, or to perform a task such as counting or singing (if they are a professional singer). Conversely, changes to the flow of speech caused by the surgical intervention are typically understood as errors or disruptions—indicators that a brain region should be avoided if normal function is to be preserved (Kim 2009). Thus, the functional roles, and corresponding mechanistic analyses, of aspects of patient speech induced through conversation and those induced through surgical intervention on the brain are quite different, and do not satisfy functional locality.[6]

This example emphasizes the importance of the empirical claim that ephaptic coupling plays a role in properly functioning neural computation. If ephaptic coupling plays a purely pathological role, as in trigeminal neuralgia, then it does not

---

[6] Plausibly, the awake craniotomy case also violates spatiotemporal locality—the Broca's areas of any two people are spatially distant, and insofar as they are engaged in a single functional interaction through conversation, the mechanism involved is transient and unstable, as is the mechanism of a surgical intervention. We do not wish to take a stand here on the importance of regular versus transient mechanisms for mechanism realism (cf. Andersen 2012), but it is worth emphasing that the ephaptic coupling case is tightly spatiotemporally circumscribed (as the token neurons involved are adjacent or at least in constant synaptic contact) unlike the case of conversation or surgery.

pose a challenge to a locally hierarchical understanding of neural computation. In that case, ephaptic coupling could indeed be understood as a process that is functionally distinct from synaptic transmission, and thus simultaneous realism about the hierarchical structure of synaptic information transmission and the hierarchical structure of ephaptic coupling would not contradict LHR, as they would represent two functionally distinct mechanism hierarchies. Recall that the target of a structural account of mechanism realism is to identify the distinctive patterns of causal and composition relations in the world that constitute mechanisms. The mere causal interaction of an EM field on a particular neuron is not by itself enough to carve up the relations that constitute mechanisms in the world, it must be subsumed under a functional analysis to count as a candidate for a real mechanism. The pathological and computational functions of ephaptic coupling would thus have different implications for mechanism realism.

5 Mechanism Structure Beyond Local Hierarchies

Mechanistic explanations are constructed through functional analysis, but functional decomposition does not guarantee that levels do not crosscut, and so inconsistent hierarchies are possible, both in principle and in practice. This result renders hierarchies of even local levels inadequate as an account of the distinctive metaphysical structure of mechanisms. The aim of this section is to formalize LHR as a first step toward specifying the space of possible forms of mechanism realism that abandon local hierarchies, yet nevertheless resist collapse into mere aggregate causes. An open question is whether weaker metaphysical positions that are consistent with crosscutting hierarchies are still 'mechanistic' in any intuitive sense.

From a structural standpoint there are three kinds of constraints needed to specify the distinctive features of mechanisms: constraints on *composition*, i.e. the relationship between entities understood at different degrees of granularity; constraints on *interaction*, i.e. the types of (typically causal) relationships that may occur between entities; and constraints on the relationship between composition and interaction. In an abstract, formal sense, both composition and interaction may be thought of as *relations* between entities, characterized by certain properties. A hierarchical mechanism may be understood as a structure in which the interplay between these two relations (compositional and causal) is mutually constraining, with causal interactions occurring only between entities at the same level of composition, and levels of composition being defined in terms of the possibility of causal interactions. The possibility of crosscutting hierarchies within a single functionally localized system shows that this condition of mutual constraint must be rejected by the mechanism realist. If this condition is rejected, however, is there any sense in which the remaining structure may be understood as distinctively *mechanistic*, as opposed to a mere aggregation of causal and compositional relations? In order to answer this question, we need to first introduce a few formal details.

Let $<$ represent the composition relation, i.e. read $a < b$ as 'a is a component of b'. As mechanists understand it, $<$ is something like the proper parthood relation studied by mereologists. Proper parthood is typically thought to be both *irreflexive* (*not*

$a < a$) and *antisymmetric* ($a < b$ and $a \neq b$ implies *not* $b < a$), i.e. $<$ is *asymmetric*: if a is a component of b, then b is not a component of a (Koslicki 2008; McDaniel 2009). This is an assumption about composition that corresponds to the strict hierarchy of layers of entities posited by LHR. Affirming that $<$ is asymmetric constitutes a substantive position, rejecting, for instance, arguments that proper parthood might be susceptible to mereological loops (Tillman and Fowler 2012). Furthermore, mereologists typically consider proper parthood to be *transitive* ($a < b$ and $b < c$ implies $a < c$), yet composition as understood in the functionally localized sense of mechanists is arguably not transitive, as (intuitively) if an entity is a component of a sub-mechanism, it is not (necessarily) also a component of the mechanism of which that sub-mechanism is a part (Seibt 2014; cf. Craver 2015: 14).[7]

Next, let $a \sqsubset b$ represent the interaction relation, i.e. 'a (causally) interacts with b'. Like $<$, one might consider $\sqsubset$ to be *irreflexive*: a does not interact with itself. Unlike $<$, $\sqsubset$ might not be anti-symmetric, though: if a and b are entities, and $a \sqsubset b$, we might also wish to allow that $b \sqsubset a$. The exact nature of $\sqsubset$ is less specified in the mechanism literature than that of $<$; nevertheless, $\sqsubset$ could be fleshed out in a variety of ways, drawing for instance on the literature on the logical structure of causality. One issue up for debate, for instance, is whether interaction should be thought of as transitive—a property endorsed by some in the logic of causation literature (Koons 1999) and denied by others (Scheffler 1993). The point is just that, for both $<$ and $\sqsubset$, there is room for debate about exactly which formal properties they will satisfy, and a full-fledged structural account of mechanism realism will need to commit on those properties.

What LHR adds to asymmetric, intransitive composition and irreflexive, (potentially) symmetric interaction is a restriction on how composition and interaction are permitted to combine, i.e. the mutual constraint that we mentioned previously. To state this restriction formally, we need a few more definitions. Write $<_T$ for the transitive closure of $<$, and $\sqsubset_{ST}$ for the symmetric, transitive closure of $\sqsubset$. The 'family' of a, F_a, is the set of all entities x such that $a <_T x$ or $x <_T a$. Crucially, note that, since composition is irreflexive, $a \notin F_a$. The "peer group" of a, $[a]$, is the set of all x such that $a \sqsubset_{ST} x$. Since $\sqsubset_{ST}$ is symmetric, $a \in [a]$. Intuitively, the family of a are all entities other than a related to it by some chain of composition relations, while the peer group is the set of all entities a is related to by some chain of interactions, i.e. all entities at the same 'level' as a. Then, the fundamental constraint on composition and interaction endorsed by LHR may be stated as:

$$\forall x \left(F_x \cap [x] = \emptyset \right)$$

This constraint rules out cross-level interactions, as required by the C–B principle. To see why, suppose a and b are on different levels of a mechanistic hierarchy, yet $a \sqsubset b$. If a and b are on different levels, then (without loss of generality) there

---

[7] Some passages in Craver (e.g. 2007:193; 2015:19) are most naturally read as advocating realism about compositional levels *only*. Without wanting to take a stance on Craver exegesis, we note that realism about compositional levels only is not a form of *mechanism* realism in our sense, as it cannot distinguish mechanisms from other aggregates of causal and compositional relations.

is some c such that $b \sqsubseteq_{ST} c$ and $c \in F_a$ (by the definition of level). But then, by the transitivity of $\sqsubseteq_{ST}$, $a \sqsubseteq_{ST} c$, and so $c \in [a]$ as well, and the constraint is violated. In other words, no member of the family of an entity may (causally) interact with it, and no member of the peer group of an entity may be in its family, or, in more familiar terms, causal interactions may only occur between entities at the same level, and entities must be at the same level if they causally interact. It is important to clarify here that x ranges only over the entities in some narrowly defined spatiotemporal region, and $\sqsubseteq$ is restricted to some set of functionally-relevant interactions, rather than over all entities and interactions in the world—this is the force of the assertion that mechanistic hierarchies are purely local. Within this narrow region, the claims of mechanism hierarchy realism are substantive, and are violated by the possibility of crosscutting functional hierarchies.

A non-trivial mechanism realism requires an account of structure in the world substantial enough to differentiate it from mere aggregate causation, yet weak enough to allow crosscutting hierarchies. What this means is that one can't simply identify as a *mechanism* realist by taking a stance on the relational structure of causation and the relational structure of composition—one must also say something about how they interact to distinguish mechanisms from bare aggregates. One could propose a weaker constraint on the relationship between composition and causation than LHR, for instance by going *very* local and defining "family" and "peer group" in terms of $<$ and $\sqsubseteq$ directly (rather than their transitive closures)—however, this still isn't going to be enough to escape crosscutting in the ephaptic coupling case, as plausibly there are no intervening compositional levels between neurons and their EM field, nor causal intermediaries between the field affected neurons.

A more promising strategy may at first seem to be simply adding additional structure. Andersen (2012), for instance, suggests a dynamical constraint on the *regularity* of activities, while Kaiser and Krickel (2017) argue that "activities" should be understood to include not only causal interactions but, say, instantiation of dispositions (amongst other additions). But simply expanding the number of ontological base categories is not by itself going to escape the problem of crosscutting hierarchies, which will continue to obtain for causality and composition no matter what additional ontological categories are added to the model. Conversely, the recent formal approach of List (2019) simply takes levels as a primitive, yet this also is unsuitable for the would-be mechanism realist, since the example of crosscutting levels shows that a hierarchy of levels may not be the correct characterization of a distinctively mechanistic organization of entities and activities.

Mechanists often state their position as if it is near trivial, feeling the need to defend it against the worry that everything constitutes a mechanism (Craver and Tabery 2015). Furthermore, they feel comfortable asserting that causal interactions or compositional relations that violate strict mechanistic hierarchies are suspect, or draped in a "shroud of mystery" (Craver and Bechtel 2007, 562). Our purpose in stating the commitments of LHR in this new way is to shift intuitions on this issue. What we hope to have accomplished in this section is to establish that (1) the assertion of hierarchy realism is substantive, and severely restricts the permitted interplay between mereology and causality, even when it is functionally and spatiotemporally localized. Furthermore, (2) while there are a variety of different choices to be

made about the formal structure of causality and composition considered separately, mechanism realism requires some further condition on the mutually constraining relationship between cause and composition, for it is precisely the *organization* of entities *and* their interactions that distinguishes mechanisms from other aggregates of entities and causes in the world. Finally, (3) while this whole discussion may be read as a reductio of the desert ontology approach, a demand for a richer set of primitives (capacities, dispositions, etc.) or more elaborate constraints (regularity) does not by itself escape the fundamental tension between functional decomposition and the embedding of mechanisms into local hierarchies.

6 Conclusion

Mechanism realists are committed to the existence of some objective structure in the world that is richer than the mere aggregation of causes. One strategy for identifying this richer mechanistic structure proceeds by functional decomposition, first identifying a system that produces a target phenomenon, and then isolating entities and their characteristic activities through the subsidiary functions they serve in generating that phenomenon. We have identified the hierarchical arrangement of mechanisms into layers of entities and activities that is presumed to fall out of functional analysis as a plausible foundation for mechanism realism. We call realism about the arrangement of entities and activities into a strict hierarchy of layers *local hierarchy realism*, to distinguish it from the *universal hierarchy realism* that is generally rejected by mechanists.

LHR has the advantage of placing a substantive and precisely stateable constraint on the relationship between mereology and causality. The organization of entities and activities to which it is committed is not reducible to mere aggregate causation, and it clearly satisfies intuitive and informal definitions of mechanism. Yet LHR is not an a priori consequence of functional decomposition. So long as mechanistic explanations are produced through functional decomposition, there is a danger they permit causal interactions that crosscut entities into competing hierarchies. Moreover, there appear to be such crosscutting hierarchies in the world, as the joint role of synaptic transmission and ephaptic coupling in the production of neural firing illustrates. So, LHR seems to be untenable.

We concluded by suggesting there are substantive positions the mechanism realist might take on the arrangement of causes and composition relations in the world that both avoid the charge that mechanisms reduce to mere aggregation of causes, yet accommodate the possibility of crosscutting hierarchies; but these weaker metaphysical commitments do not obviously satisfy intuitive notions of mechanism (as LHR did). As we see it, then, there are two paths forward for the would-be mechanism realist: (i) defend LHR by identifying a principled procedure for prioritizing hierarchies when crosscutting examples appear; or (ii) define a constraint on the relationship between mereology and causality weaker than that of LHR, yet still intuitively recognizable as satisfying informal definitions of mechanism. Alternatively, one could remain a scientific realist while abandoning mechanism realism, endorsing a

generic realism about mechanistic explanations but denying that 'mechanisms' per se are a well-defined or scientifically distinguished class of structures in the world.

References

Anastassiou, C., Perin, R., Markram, H., and Koch, C. (2011). Ephaptic coupling of cortical neurons. *Nature Neuroscience, 14,* 217–223.

Andersen, H. (2012). The case for regularity in mechanistic causal explanation. *Synthese, 189,* 415–432.

Arvanitaki, A. (1942). Effects invoked in an axon by the activity of a contiguous one. *Journal of Neurophysiology, 5*(2), 89–108.

Baumgartner, M., and Casini, L. (2017). An abductive theory of constitution. *Philosophy of Science, 84,* 214–233.

Bechtel, W. (2006). *Discovering cell mechanisms: The creation of modern cell biology*. CUP.

Bechtel, W. (2009). Looking down, around, and up: Mechanistic explanation in psychology. *Philosophical Psychology, 22*(5), 543–564.

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

Bonhomme, V., Franssen, C., and Hans, P. (2009). Awake craniotomy. *European Journal of Anaesthesiology, 26,* 906–912.

Boone, W., and Piccinini, G. (2016). The cognitive neuroscience revolution. *Synthese, 193*(5), 1509–1534.

Chiang, C.-C., Shivacharan, R., Wei, X., Gonzalez-Reyes, L., and Durand, D. (2019). Slow periodic activity in the longitudinal hippocampal slice can self-propagate non-synaptically by a mechanism consistent with ephaptic coupling. *The Journal of Physiology, 597*(1), 249–269.

Craver, C. (2001). Role functions, mechanisms and hierarchy. *Philosophy of Science, 68,* 31–55.

Craver, C. (2007). *Explaining the Brain*. OUP.

Craver, C. (2013). Functions and mechanisms: a perspectivalist view. In P. Huneman (Ed.), *Functions: selection and mechanisms* (pp. 133–158). Springer, Netherlands.

Craver, C. (2015). Levels. In: T. Metzinger, J. Windt (eds) *Open MIND* 8. Frankfurt am Main: MIND Group.

Craver, C., and Bechtel, W. (2007). Top-down causation without top-down causes. *Biology and Philosophy, 2,* 547–563.

Craver, C., J. Tabery. (2015). Mechanisms in science. In: E. Zalta (ed) The Stanford Encyclopedia of Philosophy.

Cummins, R. (1975). Functional analysis. *Journal of Philosophy, 72,* 741–764.

Darden, L. (2002). Rethinking mechanistic explanation. *Philosophy of Science, 69*(S3), 342–353.

Darden, L. (2008). Thinking again about biological mechanisms. *Philosophy of Science, 75*(5), 958–969.

Dickson, C. (2010). Ups and downs in the hippocampus. *Behavioral Brain Research, 214,* 35–41.

Elber-Dorozko, L., and Shagrir, O. (2019). Integrating computation into the mechanistic hierarchy in the cognitive and neural sciences. *Synthese*. https://doi.org/10.1007/s11229-019-02230-9

Eronen, M. (2015). Levels of organization: a deflationary account. *Biology and Philosophy, 30*(1), 39–58.

Garson, J. (2013). The functional sense of mechanism. *Philosophy of Science, 80*(3), 317–333.

Garson, J. (2017). A generalized selected effects theory of function. *Philosophy of Science, 84*(3), 523–543.

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

Glennan, S. (2002). Rethinking mechanistic explanation. *Philosophy of Science, 69*(S3), S342–S353.

Glennan, S. (2005). Modeling mechanisms. *Studies in History and Philosophy of Science C, 36*(2), 443–464.

Glennan, S. (2017). *The New Mechanical Philosophy*. OUP.

Hardcastle, V. (1999). Understanding functions: a pragmatic approach. In V. Hardcastle (Ed.), *When biology meets philosophy* (pp. 27–46). MIT Press.

Illari, P., and Williamson, J. (2012). What is a mechanism? *European Journal for the Philosophy of Science, 2*(1), 119–135.

Kaiser, M. (2007). Brain architecture: A design for natural computation. *Philosophical Transactions of the Royal Society A, 365*, 3033–3045.

Kaiser, M., and Krickel, B. (2017). The metaphysics of constitutive mechanistic phenomena. *British Journal for the Philosophy of Science, 68,* 745–779.

Kästner, L. (2018). Integrating mechanistic explanations through epistemic perspectives. *Studies in the History and Philosophy of Science, 68,* 68–79.

Kim, J. (2002). The layered model: metaphysical considerations. *Philosophical Explorations, 5*(1), 2–20.

Kim, S., et al. (2009). Awake craniotomy for brain tumors near eloquent cortex. *Neurosurgery, 64*(5), 836–846.

Koons, R. (1999). Situation mereology and the logic of causation. *Topoi, 18*(2), 167–174.

Koslicki, K. (2008). *The structure of objects*. OUP.

Krickel, B. (2018). *The mechanical world: metaphysical commitments of the mechanistic approach*. Springer.

Kundu, K., Darden, L., Moult, J. (2020). MecCog: a knowledge representation framework for genetic disease mechanism. Doi:https://doi.org/10.1101/2020.09.03.282012

List, C. (2019). Levels: descriptive, explanatory, and ontological. *Noûs, 53*(4), 852–883.

Love, S., and Coakham, H. (2001). Trigeminal neuralgia: pathology and pathogenesis. *Brain, 124*(12), 2347–2360.

Machamer, P. (2004). Activities and causation: the metaphysics and epistemology of mechanisms. *International Studies in the Philosophy of Science, 18*(1), 27–39.

Machamer, P., Darden, L., and Craver, C. (2000). Thinking about mechanisms. *Philosophy of Science, 67*(1), 1–25.

Maley, C., and Piccinini, G. (2017). A unified mechanistic account of teleological functions for psychology and neuroscience. In D. Kaplan (Ed.), *Explanation and integration in mind and brain science* (pp. 236–256). OUP.

McDaniel, K. (2009). Extended simples and qualitative heterogeneity. *The Philosophical Quarterly, 59,* 325–331.

Millikan, R. (1989). In defense of proper functions. *Philosophy of Science, 56,* 288–302.

Neander, K. (1991). Functions as selected effects. *Philosophy of Science, 58*(2), 168–184.

Persson, J. (2010). Activity-based accounts of mechanism and the threat of polygenic effects. *Erkenntnis, 72*(1), 135–149.

Piccinini, G., and Bahar, S. (2013). Neural computation and the computational theory of cognition. *Cognitive Science, 34,* 453–488.

Piccinini, G., and Craver, C. (2011). Integrating psychology and neuroscience: functional analyses as mechanism sketches. *Synthese, 183*(3), 283–311.

Potochnik, A. (2017). *Idealization and the aims of science*. University of Chicago Press.

Povich, M., and Craver, C. (2017). Mechanistic levels, reduction, and emergence. In I. Glennan (Ed.), *The Routledge handbook of mechanisms and mechanical philosophy*. Routledge.

Psillos, S. (2004). A glimpse of the *Secret Connexion*: harmonizing mechanisms with counterfactuals. *Perspectives on Science, 12*(3), 288–319.

Putnam, H., Oppenheim, P. (1958). Unity of science as a working hypothesis. In: Feigl et al. (eds.), *Minnesota studies in the philosophy of science, vol. 2*. Minneapolis, MN: Minnesota University Press.

Raerinne, J. (2011). Causal and mechanistic explanations in ecology. *Acta Biotheoretica, 59*(3–4), 251–271.

Romero, F. (2015). Why there isn't inter-level causation in mechanisms. *Synthese, 192,* 3731–3755.

Scerrati, A., et al. (2020). Artists playing music while undergoing brain surgery: a look into the T scientific evidence and the social media perspective. *Clinical Neurology and Neurosurgery., 196,* 105911.

Scheffler, U. (1993). On the logic of event causation. *Logic and Logical Philosophy, 1,* 129–155.

Seibt, J. (2014). Non-transitive parthood, leveled mereology, and the representation of emergent parts of processes. *Grazer Philosophische Studien., 91,* 165–190.

Shagrir, O., and Bechtel, W. (2017). Marr's computational level and delineating phenomena. In D. Kaplan (Ed.), *Explanation and integration in mind and brain science* (pp. 190–214). OUP.

Shepherd, G. M. (1994). *Neurobiology* (3rd ed.). OUP.

Sheredos, B., Burnston, D., Abrahamsen, A., and Bechtel, W. (2013). Why do biologists use so many diagrams? *Philosophy of Science, 80*(5), 931–944.

Tillman, C., and Fowler, G. (2012). Propositions and parthood: the universe and anti-symmetry. *Australasian Journal of Philosophy, 90*(3), 525–539.

Wilhelm, I. (2019). The ontology of mechanisms. *The Journal of Philosophy, 116*(11), 615–636.