Decoupling topological explanations from mechanisms

Kostic, D.; Khalifa, K.

Philosophy of Science (2023), **90**, 245–268
doi:10.1017/psa.2022.29

ARTICLE

Decoupling Topological Explanations from Mechanisms

Daniel Kostić[1]* and Kareem Khalifa[2]

[1]Radboud Excellence Initiative Fellow, Institute for Science in Society (ISiS), Radboud University, Nijmegen, The Netherlands and [2]Department of Philosophy, Middlebury College, Middlebury, VT, USA
**Corresponding author*. Emails: daniel.kostic@gmail.com, kkhalifa@middlebury.edu

(Received 20 October 2020; revised 30 November 2021; accepted 01 April 2022; first published online 28 April 2022)

Abstract

We provide three innovations to recent debates about whether topological or "network" explanations are a species of mechanistic explanation. First, we more precisely characterize the requirement that all topological explanations are mechanistic explanations and show scientific practice to belie such a requirement. Second, we provide an account that unifies mechanistic and non-mechanistic topological explanations, thereby enriching both the mechanist and autonomist programs by highlighting when and where topological explanations are mechanistic. Third, we defend this view against some powerful mechanist objections. We conclude from this that topological explanations are autonomous from their mechanistic counterparts.

1. Introduction

Recent literature discusses topological or "network" explanations in the life sciences (Bechtel 2020; Craver 2016; Darrason 2018; Green et al. 2018; Huneman 2010; Jones 2014; Kostić 2018, 2019, 2020; Kostić and Khalifa 2021; Levy and Bechtel 2013; Matthiessen 2017; Rathkopf 2018; Ross 2020). Some self-described mechanists have only focused on topological explanations as a species of mechanistic explanation (Bechtel 2020; Craver 2016; DiFrisco and Jaeger 2019; Levy and Bechtel 2013). For them, topological explanations appear to rely on the time-honored scientific strategy of understanding a system by grasping how its components interact with each other. However, others—whom we shall dub "(topological) autonomists"—claim that topological explanations are markedly different than their mechanistic counterparts (Darrason 2018; Huneman 2010, 2018a; Kostić 2018, 2019, 2020; Rathkopf 2018).

These debates have far-ranging implications. Ideally, they would provide scientists with guidelines for when mechanistic information is a prerequisite for a topological model's being explanatory. Furthermore, if autonomists are correct, the debates should provide additional guidelines for when mechanistic information is simply an added bonus to a free-standing topological explanation. In addition to their scientific upshots, these debates between mechanists and autonomists contribute to wider philosophical discussions concerning explanatory pluralism, noncausal explanation, modeling, and the applicability of mathematics. Finally, engagement with topological explanations enriches both the mechanist and autonomist programs by highlighting when and where topological explanations are mechanistic.

To make good on these promises, both mechanists and autonomists would benefit substantially from a more precise and systematic account of when topological explanations count as mechanistic. What kinds of considerations would either differentiate or unify topological and mechanistic explanations? This paper aims to fill this gap. The result is a novel argument for autonomism. To that end, Section 2 describes topological explanations in relatively neutral terms. Section 3 then presents and motivates, in our estimate, the most plausible and principled way of interpreting the claim that topological models are explanatory only insofar as they are mechanistic. Section 4 then provides a neuroscientific example that does not fit this mechanistic framework. Section 5 provides an autonomist alternative to the account of topological explanation offered in Section 3, which unifies topological explanations of both the mechanistic and non-mechanistic varieties. Section 6 then shows how this account of topological explanation rebuts some powerful objections to autonomism.

2. Background

Before embarking on these philosophical tasks, we review topological explanations' basic concepts. Topological explanations describe how their respective explananda depend upon topological properties. For the purposes of this essay, we will focus on those topological properties that can be represented using the resources of graph theory.[1] A graph is an ordered pair (V, E), where V is a set of *vertices* (or nodes) and E is a set of *edges* (links, or connections) that connect those vertices. For ease of locution, we will use the term "graph" or "topological model" to denote the mathematical representation of a topological structure, and "network" to denote a real-world structure (van den Heuvel and Sporns 2013, 683).

Vertices and edges represent different things in different scientific fields. For example, in neuroscience, vertices frequently represent neurons or brain regions; while edges represent synapses or functional connections. In computer science, graphs frequently represent networks of cables between computers and routers or networks of hyperlinked web pages. In ecological and food networks, vertices might be species; edges, predation relations.

Scientists infer a network's structure from data, and then apply various graph-theoretic algorithms to measure its topological properties. For instance, clustering coefficients measure degrees of interconnectedness among nodes in the same neighborhood. Here, a node's *neighborhood* is defined as the set of nodes to which it is directly connected. An individual node's *local* clustering coefficient is the proportion of edges within its neighborhood divided by the number of edges that could possibly exist between the members of its neighborhood. By contrast, a network's *global*

[1] Since the mechanist-autonomist debates chiefly concern network topology in the graph-theoretic sense, we will use "topology" and "graph-theory" interchangeably. Consequently, we bracket non–graph theoretic branches of topology.

https://doi.org/10.1017/psa.2022.29 Published online by Cambridge University Press

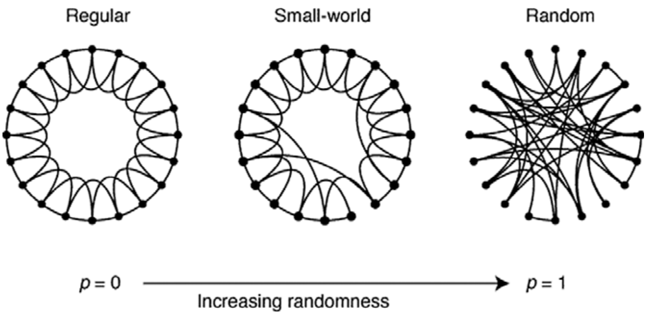

Figure 1. Different kinds of networks discussed in Watts and Strogatz (1998).

clustering coefficient is the ratio of closed triplets to the total number of triplets in a graph. A triplet of nodes is any three nodes that are connected by at least two edges. An *open* triplet is connected by exactly two edges; a *closed* triplet, by three. Another topological property, average (or "characteristic") path length, measures the mean number of edges needed to connect any two nodes in the network.

In their seminal paper, Watts and Strogatz (1998) applied these concepts to show how a network's topological structure determines its dynamics. First, *regular graphs* have both high global clustering coefficients and high average path length. By contrast, *random graphs* have low global clustering coefficients and low average path length. Finally, they introduced a third type of *small-world graph* with high clustering coefficients but low average path length (Figure 1).

Highlighting differences between these three types of graphs yields a powerful explanatory strategy. For example, because regular networks have larger average path lengths than small-world networks, things will "spread" throughout the former more slowly than the latter, largely due to the greater number of edges to be traversed. Similarly, because random networks have smaller clustering coefficients than small-world networks, things will also spread throughout the former more slowly than the latter, largely due to sparse interconnections within neighborhoods of nodes. Hence, ceteris paribus, propagation is faster in small-world networks. This is because the fewer long-range connections between highly interconnected neighborhoods of nodes shorten the distance between neighborhoods of nodes that are otherwise very distant, which enables them to behave as if they were first neighbors. For example, Watts and Strogatz showed that the nervous system of *C. elegans* is a small-world network, and subsequent researchers argued that this system's small-world topology explains its relatively efficient information propagation (Bullmore and Sporns 2012; Latora and Marchiori 2001).

3. Mechanism and topology

Are topological explanations such as the one involving *C. elegans* just mechanistic explanations in fancy mathematical clothing? To answer this question, we first clarify what we mean by "mechanistic explanation" (Section 3.1), and then present a framework for interpreting topological explanations mechanistically (Section 3.2). This provides the stiffest challenge to autonomism, and thereby sets the stage for Section 4,

https://doi.org/10.1017/psa.2022.29 Published online by Cambridge University Press

where we show that autonomism is nevertheless unfazed by this mechanistic contender.[2]

3.1. Mechanistic explanation

Before evaluating whether topological explanations are mechanistic explanations, we elucidate the latter. For our purposes, we focus on conceptions of mechanistic explanation that are minimal and nontrivial. Such conceptions provide the most formidable challenges to autonomism. Hence, when we show that some topological explanations are not mechanistic, we cannot be charged with chasing ghosts. We discuss these two facets of mechanistic explanation in turn.

To begin, Glennan (2017, 17) provides a "minimal" characterization of mechanisms that captures a widely held consensus among mechanists about conditions that are necessary for something to be a mechanism, even if they differ about, for example, the role of regularities, counterfactuals, and functions in mechanistic explanation:

> A mechanism for a phenomenon consists of entities (or parts) whose activities and interactions are organized so as to be responsible for the phenomenon.[3]

Because this conception is *minimal*, the counterexamples we raise to it below apply a fortiori to more demanding conceptions of mechanisms.

Importantly, mechanists are sometimes criticized for being so vague as to *trivialize* the concept of mechanism (Dupré 2015). Since minimal conceptions are less committal than other conceptions of mechanism, they are especially susceptible to this criticism. To avoid trivializing mechanistic explanation, we appeal to claims about "entities," "activities," "interactions," "organization," and "responsibility" widely endorsed by mechanists. For instance, our arguments hinge on claims that entities and activities cannot be spatiotemporal regions determined merely by convention; that interactions cannot be mere correlations, etc. We further develop these claims as they arise in the discussion of our examples of non-mechanistic topological explanation. Should mechanists deny these requirements on entities, activities, and the like, then the burden of proof falls upon them to show that their alternative conception of mechanism is nontrivial.

Finally, mechanists typically distinguish between *etiological*, *constitutive,* and *contextual* mechanistic explanations (e.g., Craver 2001). Etiological explanations cite the causal history of the explanandum; constitutive explanations cite the underlying mechanism of that explanandum; and contextual explanations cite an explanandum's contribution to the mechanism of which it is a part. Since underlying and overlying mechanisms can be synchronic with the phenomena that they explain, neither constitutive nor contextual explanations must cite their respective explananda's causal histories, i.e., the prior

[2] The ideas developed in Sections 3.1 and 3.2 are offered as a new foil to autonomism; they are not intended to systematize earlier criticisms of autonomism. Indeed, apart from a few mechanists discussed in Section 4.1 and Craver (2016), whom we discuss at length in Section 6, other challenges to autonomism are orthogonal to the position we develop below (Bechtel 2020; Green et al. 2018; DiFrisco and Jaeger 2019; Matthiessen 2017; Ross 2020).

[3] Craver and Tabery (2019) and Illari and Williamson (2010) provide similar "minimal" characterizations.

https://doi.org/10.1017/psa.2022.29 Published online by Cambridge University Press

events that produced the phenomenon. Therefore, constitutive and contextual explanations are distinct from etiological explanations. Furthermore, because constitutive explanations explain a larger system in terms of its parts, while contextual mechanisms do the exact opposite, constitutive and contextual explanations are also distinct. For ease of exposition, we will first discuss constitutive mechanistic explanations. Section 4.4 discusses their etiological and contextual counterparts.

To summarize, we are trying to precisely characterize mechanistic explanations in a way that poses the stiffest challenge to our claim that some topological explanations are non-mechanistic. To that end, we think that the most defensible conception of mechanistic explanation has two features: minimality and nontriviality. Furthermore, we will first compare topological explanations to constitutive mechanistic explanations, and then turn to etiological and contextual ones.

3.2. Mechanistic interpretations of topological explanations

With a clearer conception of mechanistic explanation in hand, mechanists' next task is to translate topological explanations' characteristic graph-theoretic vocabulary into the language of mechanistic explanation—to provide a mechanistic interpretation of topological explanations (MITE). To that end, the preceding suggests that mechanists ought to hold that a topological model is explanatory only insofar as there exists a mechanism for which all the following conditions hold:

(1) *Node Requirement*: the topological model's nodes denote the mechanism's entities or activities.
(2) *Edge Requirement*: the topological model's edges denote the interactions between the mechanism's entities or activities.
(3) *Responsibility Requirement*: the topological model specifies how the mechanism's entities, activities, and interactions are organized[4] so as to be responsible for the phenomenon.
(4) *Interlevel Requirement*: the explanandum is at a higher level than the mechanism's entities, activities, and interactions as described by the Node and Edge Requirements.

The first three requirements fall out of Glennan's minimal conception of mechanism; the last, from our initial focus on constitutive explanations.

As an illustration of the MITE's plausibility, consider the earlier example in which the small-world topology of the *C. elegans* nervous system explains its capacity to process information more efficiently than would be expected if its topology were either regular or random. Here, the neural system's components are individual neurons, which are represented as nodes in the topological model. Hence, the Node Requirement is satisfied. Furthermore, the edges of the graph denote synapses or gap junctions between different neurons, so the Edge Requirement is satisfied.

[4] Given the definition of a graph, the Node and Edge Requirements entail an "Organization Requirement," according to which the topological model must specify how entities or activities are organized with respect to their interactions. Because the Organization Requirement is a mere consequence of the Node and Edge Requirements, we keep it implicit hereafter, though we discuss some mechanist proposals for organization below.

Third, the Responsibility Requirement is satisfied, though this requires more detailed discussion. Suppose (as we shall throughout) that "responsibility" is characterized in terms of counterfactual dependence:

> Had the *C. elegans* neural network been regular or random (rather than small-world), then information transfer would have been less efficient (rather than its actual level of efficiency).

In regular or random topologies, the synaptic connections between neurons will be different than they are in the actual small-world topology exhibited in *C. elegans*. Consequently, each of these networks describes a different potential mechanistic structure. Thus, this counterfactual shows how mechanistic differences are responsible for differences in information transfer. Finally, note that the efficiency of information transfer is a global property of the *C. elegans* nervous system. That system is composed of neurons connected via synapses, so mechanists appear justified in claiming that the nodes and edges of this network are at a lower level than its explanandum property. Hence, the Interlevel Requirement is satisfied. Putting this all together, this means that the example fits the MITE.[5]

Before proceeding, we note three things. First, while other MITEs are certainly possible, this one strikes us as the most plausible. We have culled it from some of the foremost mechanists' discussions of topological explanations (Craver 2016; Levy and Bechtel 2013, Glennan 2017). Indeed, our MITE also accords nicely with the widely used "Craver diagrams" in the mechanisms literature. As Figure 2 illustrates, such diagrams entail that there is a phenomenon ("*S's* Ψing") at a higher level, and a mechanism exhibiting a graph-theoretic structure at the lower level, with nodes corresponding to entities (denoted by "X_i") performing activities (denoted by "ϕ_i"), and edges corresponding to interactions (denoted by arrows). Since no other MITE has been offered, mechanists ought to propose an alternative MITE should they chafe at the one we propose here. Second, our working definition of autonomism throughout this paper is only that *some* topological explanations do *not* fit this MITE. This is consistent with *some other* topological explanations fitting this MITE. Third, it suffices for our purposes if only one of the MITE's four requirements is violated. In other words, so long as our counterexample is even "partly non-mechanistic", autonomism (*sub specie* this MITE) is vindicated.

4. Non-mechanistic topological explanation

Autonomists have distanced themselves from mechanistic explanations in myriad ways. For instance, Rathkopf (2018) argues that topological explanations are normally used in nearly decomposable and nondecomposable systems, whereas mechanistic explanations are typically used in decomposable systems. Another autonomist approach treats topological explanations as conferring mathematical necessity upon their explananda (Huneman 2018a; Lange 2017). Still others claim that topological explanations are frequently more abstract than mechanistic explanations

[5] Anyone who denies that this is sufficient for establishing that the explanation is mechanistic already subscribes to autonomism or should offer an alternative MITE.

https://doi.org/10.1017/psa.2022.29 Published online by Cambridge University Press

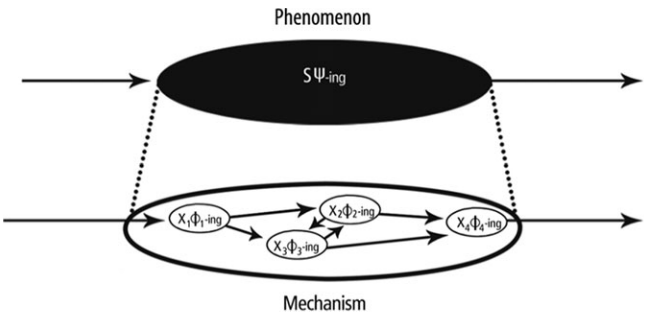

Figure 2. Craver (2007) diagram.

(Darrason 2018; Huneman 2018a; Kostić 2019). Ross (2020) and Woodward (2013) suggest that some topological explanations are resistant to the kinds of interventions that are characteristic of causal-mechanical explanations. Finally, Huneman (2018b) and Kostić (2018) argue that topological and mechanistic explanations involve different kinds of realization relations.

We provide a new argument for autonomism. So far as we can tell, it complements rather than competes with these other autonomist arguments, and has the added virtue of engaging a more precise mechanistic foil—the MITE developed above. Specifically, we use this foil to provide an example of a non-mechanistic topological explanation: Adachi et al.'s (2011) explanation of the contribution of anatomically unconnected areas to functional connectivity in macaque neocortices. This explanation rests on the neuroscientific distinction between *anatomical connectivity* (AC) and *functional connectivity* (FC). While both kinds of connectivity are modeled using graph theory, only AC networks are naturally glossed as mechanisms. For instance, nodes are segregated anatomical regions of the brain (e.g., different Brodmann areas, gyri, and cortical lobes) and their edges are causal relations (what is sometimes called "effective connectivity") that are frequently identified with axonal signal flows. Since this is the topological model that figures in Adachi et al.'s explanans, we will grant that it satisfies the MITE's Node and Edge Requirements. However, we deny that this is a mechanistic explanation, for it violates the MITE's Responsibility and Interlevel Requirements. After describing the explanation in some detail (Section 4.1), we examine these violations in turn (Sections 4.2 and 4.3). This shows that this explanation is not a constitutive mechanistic explanation. We round out our defense of autonomism by anticipating and rebutting two possible mechanist responses to our argument (Sections 4.4 and 4.5).

4.1. Adachi et al.'s explanation

We first discuss Adachi et al.'s explanandum and then their explanans. Adachi et al.'s dependent variable is (combined) ΔFC, "the regression slope of FC on the total number of length2-AC" patterns.[6] Obviously, a better understanding of our explanandum, ΔFC, requires an understanding of both FC and length2-AC.

[6] Strictly speaking, the researchers explain two related phenomena. First, they claim that the frequency of both two-node and three-node motifs in the AC network explain why a *particular* length2-AC pattern *j*, which may be *a*, *b*, or *c*, affects FC to the degree that it does ("ΔFC_j"). Second, they claim

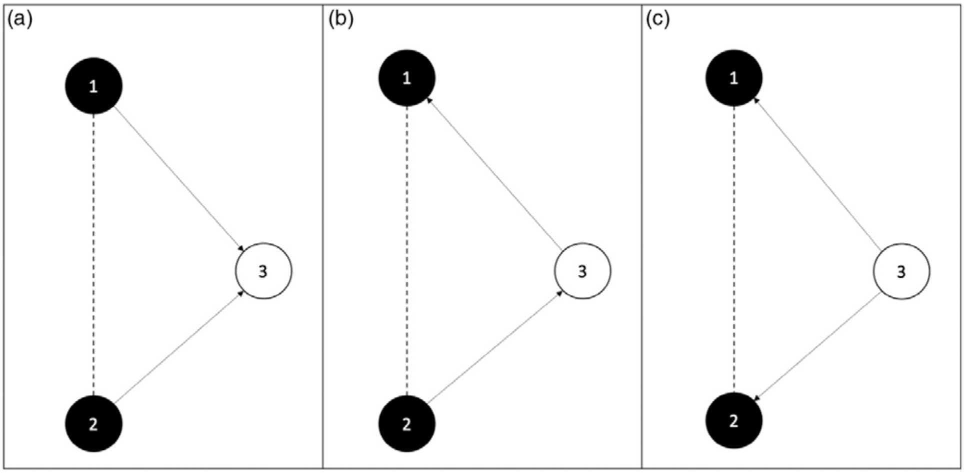

Figure 3. Different anatomical connectivity patterns (length = 2) discussed by Adachi et al. (2011). The black nodes, 1 and 2, are functionally connected by the dotted line, but are only indirectly anatomically connected via node 3 by the arrows.

Begin with FC. FC networks' edges are synchronization likelihoods (SL). Stam et al. (2006, 93) provide a useful definition:

> The SL is a general measure of the correlation or synchronization between 2 time series . . . . The SL is then the chance that pattern recurrence in time series X coincides with pattern recurrence in time series Y.

As this definition suggests, FC networks' nodes are time series. Depending on the study, these nodes can be interpreted in multiple ways. Many FC models are interpreted so that nodes correspond to the time series of blood oxygen level-dependent (BOLD) readings for an individual voxel in functional magnetic resonance imaging (fMRI) data. A voxel is a unit of graphic information that defines a three-dimensional region in space. Voxels are essentially the result of dividing the brain region of interest into a three-dimensional grid. By contrast, Adachi et al. (2011, 1587) opt for a less "operational" interpretation of their FC model. They do this by mapping different clusters of voxels onto 39 different anatomical regions or "areas" in the macaque cortex. Examples include visual area 4 and the mediodorsal parietal area. Consequently, their FC and AC models have the same nodes.

Indeed, without this mapping of FC nodes onto anatomical regions, the neuroscientists could not characterize their explanandum with any precision. Specifically, they focus on "length2-AC" patterns—i.e., functionally connected pairs of regions that are only anatomically connected through some third region (see Figure 3). In other words, length2-AC patterns involve two brain areas that are

that the frequency of three-node motifs explains why the totality of these patterns contribute to FC to the degree that they do ("combined ΔFC"). Our discussion focuses on the latter, though our arguments apply to the former as well.

https://doi.org/10.1017/psa.2022.29 Published online by Cambridge University Press

functionally connected (and hence are correlated) but are also known to lack any direct causal link. Hence, the explanandum, ΔFC, describes the extent to which the totality of Patterns *a, b,* and *c* contributes to the overall FC in the macaque neocortex.

With the explanandum clarified, we turn to Adachi's explanans, which appeals to the global topological properties of the AC network. Specifically, Adachi et al. argue that the AC network's frequency of three-node motifs (which they abbreviate as "MF3") explains why the macaque's ΔFC is as high as it is. In this context, a three-node motif is a triplet where the nodes denote anatomical regions and the edges denote anatomical connections. Thus, MF3 is a measure of how many of these triplets can be found in the macaque neocortex.

Adachi et al. establish this explanation by running simulations involving thousands of randomly generated networks. Some of these networks have the same frequency of three-node motifs as the macaque brain but differ with respect to other topological properties. These other properties include global clustering coefficient, modularity, and the frequency of two-node motifs. Then, each network's ΔFC is measured. In another run of simulations, they controlled for these other topological properties while varying the frequency of three-node motifs. Models in which the frequency of three-node motifs matched the macaque neocortex vastly outperformed models matching other topological properties in accounting for ΔFC. Once again using counterfactual dependence as a working definition of "responsibility," this suggests the following:

> Had macaque neocortices had a different frequency of three-node network motifs (rather than a different clustering coefficient, modularity, or frequency of two-node network motifs), then these neocortices' ΔFC would have been different (rather than its actual amount).

In other words, MF3 is "responsible" for the length2-AC patterns' contributions to FC.

4.2. Responsibility requirement.[7]

We now turn to our broader aim of arguing for autonomism, starting with the Responsibility Requirement, which states that topological models are only explanatory if they specify how the phenomenon counterfactually depends on how a mechanism's entities, activities, and interactions are organized. We shall consider different mechanist proposals for how Adachi et al.'s explanation satisfies the Responsibility Requirement and show that each faces serious challenges.

The most prominent mechanist strategy holds that many topological explanations describe an abstract kind of mechanistic organization (e.g., Bechtel 2009; Glennan 2017; Kuorikoski and Ylikoski 2013; Levy and Bechtel 2013; Matthiessen 2017). This "organization strategy" prompts three replies. First, the most precise versions of this

[7] Admittedly, we have made some nontrivial assumptions throughout the rest of Section 4 to even get this explanation to *appear* mechanistic. Should those assumptions be incongruous with Adachi et al.'s reasoning, then the burden of proof falls on mechanists to provide an alternative interpretation. We note that even with these generous allowances, Adachi et al.'s explanation does not fit the MITE.

https://doi.org/10.1017/psa.2022.29 Published online by Cambridge University Press

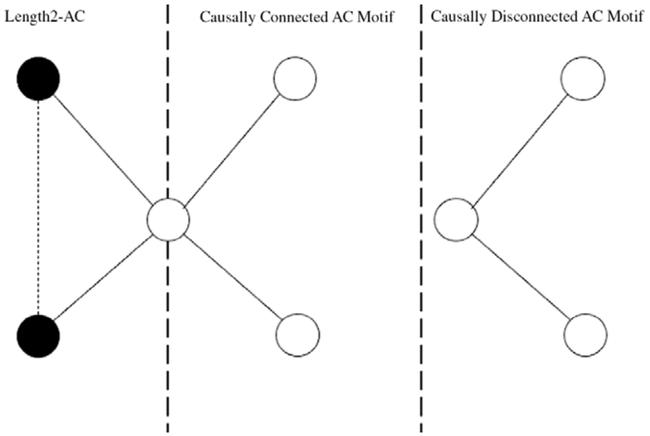

Figure 4. Examples of three kinds of three-node anatomical connectivity motifs in Adachi et al. (2011). The dotted line in between the black nodes in the length2-AC pattern is a functional connection. The other, heavier dotted lines simply distinguish the three kinds of three-node anatomical motifs. Solid lines denote anatomical connections; arrows have been removed to indicate neutrality about the direction of causation between different anatomical regions.

strategy simply *assume* that mechanistic organization can be spelled out *entirely* in terms of graph-theory. This effectively concedes autonomism's main tenets. For instance, Kuorikoski and Ylikoski (2013) *define* "organization" and "network structure" in terms of topological structure, as evidenced by both their examples and their acknowledgement of Watts as pioneering the "new science of networks." However, this means that mechanistic organization *just is* topological structure. Since they also hold that some organization is explanatory unto itself, their position entails that some topological properties are explanatory unto themselves. This is tantamount to autonomism. Second, organization is supposed to refer to the structure of interactions between entities and activities that are constitutive *of* the phenomenon to be explained. Yet, in this example, the organization in question is of a system that is (partly) constituted *by* the phenomenon to be explained. After all, the anatomical regions that figure in each of these length2-AC patterns are parts of the anatomical network, and the latter's frequency of three-node motifs drives the explanation. Third, different models with the same frequency of three-node motifs can posit radically different interactions between any three brain regions—radically different forms of organization—yet nevertheless predict the same FC structure.[8] This would mean that the specific interactions between the mechanistic components are explanatorily idle, which also violates the Responsibility Requirement.

Chastened by these problems, a mechanist might remain silent about organization, and instead insist that Adachi et al.'s explanation satisfies the Responsibility

[8] Levy and Bechtel (2013) discuss motifs' explanatory significance in this way. However, unlike Adachi et al.'s explanation, they discuss examples in which the explanans is not the mere *frequency* of motifs, which is what drives our non-mechanistic interpretation of this example.

https://doi.org/10.1017/psa.2022.29 Published online by Cambridge University Press

Requirement by specifying how ΔFC counterfactually depends on the macaque neocortex's entities, activities, and interactions. However, this faces something we call the *causal disconnection problem.* In a nutshell, the problem is that: (a) mechanistic explanations require entities, activities, and interactions that are responsible for a phenomenon to be causally connected to that phenomenon,[9] yet, (b) Adachi et al.'s explanation does not require these causal connections. To see why, let us distinguish different kinds of three-node anatomical motifs included in their calculation of MF3 (see Figure 4). If Adachi et al.'s explanation is mechanistic, then the only three-node anatomical motifs that can figure in Adachi et al.'s explanation are length2-AC patterns[10] and anatomical motifs that are causally connected to a length2-AC pattern. However, Adachi et al.'s model implies that even if the only change in MF3 were to the number of three-node motifs that are causally *disconnected* from length2-AC patterns, ΔFC would still change. Consequently, this explanation does not satisfy the MITE's Responsibility Requirement.

Finally, mechanists might try to avoid these problems by insisting that causally disconnected three-node AC motifs are explanatorily irrelevant and only the length2-AC patterns and the AC-motifs that are causally connected to them are responsible for ΔFC. However, this proposal faces two challenges. First, mechanists would need a non-question-begging argument as to why models such as Adachi et al.'s are incorrect to treat causally disconnected motifs as explanatorily relevant. To our knowledge, no such arguments have been offered. Second, the Responsibility Requirement is not easily satisfied even if causally disconnected three-node AC motifs are excluded from the explanation. Indeed, Adachi et al. (2011, 1589) turn to a non-mechanistic topological explanation precisely because of limitations in causal explanations of a phenomenon closely related ΔFC: why there is *any* functional connectivity in a length2-AC pattern. On the current mechanistic construal, length2-AC patterns' functional connections (i.e., the edge between nodes 1 and 2 in each of the patterns in Figure 3) should be explained by some indirect causal link; paradigmatically, via a third region serving as either an intermediate cause or a common cause of the functional connection between the two regions (Patterns *b* and *c* in Figure 3.) However, the anatomical structure in Pattern *b*—the "two-step serial relay"—*decreases* FC. More importantly, Adachi et al. observe that a significant number of functional connections occur even when two functionally connected areas only share a common effect (specifically: a common efferent as represented by Pattern *a*.) To our knowledge, no mechanists claim that an effect can be responsible for its cause. Hence, functionally connected areas that are only anatomically connected via a common efferent (as in Pattern *a*) cannot satisfy the Responsibility Requirement. Consequently, even if the causal disconnection problem is bracketed, mechanisms alone cannot be responsible for ΔFC, which (to repeat) is the *totality* of functional connectivity for which length2-AC patterns are responsible.

In summary, we have considered three ways of trying to mechanistically interpret MF3: as describing a mechanism's organization; as describing a mechanism's entities,

[9] At the very least, the burden of proof would be on mechanists to argue otherwise without trivializing the concept of mechanism. Our arguments in this section aim to show that this endeavor is no small task.

[10] Length2-AC patterns are three-node anatomical motifs. To see this, simply remove the functional connection (dotted line) from each of the length2-AC patterns in Figure 3.

https://doi.org/10.1017/psa.2022.29 Published online by Cambridge University Press

activities, and interactions while allowing for causal disconnections; and as describing a mechanism's entities, activities, and interactions while prohibiting causal disconnections. In each case, the claim that a mechanism is responsible for ΔFC faces formidable challenges.

4.3. Interlevel requirement

Adachi et al.'s explanation also violates the Interlevel Requirement. We provide three arguments. First, constitutive mechanistic explanations require explananda to be at higher levels than the parts and activities that explain them. However, given the way that Adachi et al. interpret their FC model, both the FC and AC networks have the same brain regions as their nodes. Because the explanans and explanandum appeal to the same entities, the FC network is not at a higher level than the AC network in this explanation.[11] Moreover, the anatomical regions that figure in MF3 are minimally at the same level as the anatomical regions in the length2-AC patterns that figure in ΔFC. Indeed, the explanation allows for some anatomical regions that figure in MF3 to be *identical* to those that figure in ΔFC.

In our estimate, this first argument understates the degree to which this explanation violates the Interlevel Requirement. This leads to our second argument. As already noted, a constitutive mechanistic explanation of ΔFC would need to appeal to the entities, activities, and interactions constitutive *of* (and thus at a lower level than) the anatomical regions and their axonal connections—and this explanation works in the exact *opposite* direction. Adachi et al. appeal to the frequency of three-node network motifs, a global property of the AC network. This network is constituted *by* these anatomical regions. In other words, a higher-level property is explaining the behavior of lower-level constituents. Hence, the MITE's Interlevel Requirement has been violated.[12]

Finally, one may argue that both ΔFC and MF3 are global properties of the macaque neocortex. If this is correct, then it is not even clear that Adachi et al.'s explanation appeals to levels at all.

4.4. Contextual and etiological explanation

This explanation's violation of the Interlevel Requirement may prompt mechanists to reply that this merely shows it not to be a *constitutive* mechanistic explanation. However, it may still be either a *contextual* or *etiological* mechanistic explanation. We briefly show that such replies face significant challenges.

Craver (2001, 63) provides the most precise definition of contextual mechanistic explanation:

> A *contextual description* of some *X*'s ϕ-ing characterizes its mechanistic role; it describes X (and its ϕ-ing) in terms of its contribution to a higher (+ 1) level

[11] For related arguments, see Kostić (2019) on the "immediacy" and Rathkopf (2018) on the "near decomposability" of some topological explanations.

[12] For related arguments, see Huneman (2018b) and Kostić (2018) on realization in topological explanations.

https://doi.org/10.1017/psa.2022.29 Published online by Cambridge University Press

> mechanism. The description includes reference not just to *X* (and its ϕ-ing) but also to *X*'s place in the organization of *S*'s ψ-ing.

If Adachi et al.'s explanation is a contextual mechanistic one, then the system (*S*) is the macaque's neocortex, the relevant system capacity (ψ) is its frequency of three-node motifs, the component *X*'s are the length2-AC patterns in the neocortex, and the activity/property (ϕ) of these patterns is their functional connectivity.

In Craver's definition, a part's activities are supposed to "contribute" to the system's capacities. We take this to require, at a minimum, that *S*'s ψ-ing *counterfactually depends* on *X*'s ϕ-ing, i.e.,

> Had it not been the case that *X* ϕ's, then it would not have been the case that S ψ's.

This accords well with Craver's paradigmatic examples of contextual mechanistic explanations, e.g., the heart pumping blood.[13] This contributes to circulation and, consonant with our suggestion, the following is true:

> Had the heart not pumped blood, then it would not have distributed oxygen and calories throughout the body.

By analogy, if Adachi et al.'s explanation is contextually mechanistic, then the following should figure prominently:

> Had macaque neocortices' ΔFC been different (rather than its actual amount), then these neocortices would have had a different frequency of three-node network motifs (rather than their actual frequency).

However, a quick review of Section 4.1 shows this to be precisely the *converse* of the counterfactual that figures in Adachi et al.'s explanation. Nor does this counterfactual accord with their methodology: They run simulations in which they vary MF3 to see how ΔFC changes—not the other way around. Consequently, ΔFC does not contribute to MF3; this is not a contextual mechanistic explanation.

As an alternative, mechanists might claim that Adachi et al.'s explanation is an *etiological* mechanistic explanation. Here, Bechtel's (2009, 557–59) account of "situated mechanisms" (or mechanistic explanations that "look up") seems especially instructive. Bechtel's chief example is how stimuli in the environment contribute to the mechanistic explanation of visual processing. Quite plausibly, environmental stimuli provide *etiological* explanations of why the mechanism for vision behaves as it does. For example, an apple and various environmental conditions (lighting, the absence of smoke and mirrors, etc.) are part of the causal history of why a person comes to see an apple. So, by analogy, if Adachi et al.'s explanation involves a situated mechanism, then MF3 is a "network environment" that is part of ΔFC's causal history.

Like Craver's account of contextual mechanistic explanation, situated mechanisms appeal to a higher-level mechanism or environment to explain a lower-level entity.

[13] It works well with Craver's other examples as well, e.g., neurotransmitters. Our arguments could be illustrated with these examples as well.

https://doi.org/10.1017/psa.2022.29 Published online by Cambridge University Press

However, unlike contextual mechanisms, situated mechanistic explanations do not require the lower-level entity to "contribute" to a higher-level mechanism in the ways we have discussed. For instance, a visual system does not need to contribute to the lighting conditions in its environment. This immediately avoids the problems raised by treating Adachi et al.'s explanation as contextually mechanistic.

Despite these initial attractions for the mechanist, other autonomists have raised challenges with identifying topological and etiological mechanistic explanations that apply here. For instance, while causes precede their effects, topological explanantia need not precede topological explananda (Huneman 2010, 218–19). Since MF3 does not temporally precede ΔFC, the explanation is not etiological.[14] Thus, MF3 is not an "environmental condition" in Bechtel's sense.[15]

4.5. Defending explanatoriness

At this point, we have at least shown that Adachi et al.'s topological model is not mechanistic. Assuming our arguments are sound, mechanists' only recourse is to deny that this model is explanatory. So far as we can tell, the most plausible argument to this effect is best glossed as a *methodological* concern about the design of the study: Do the statistical and computational models provide sufficient *evidence* for ΔFC's counterfactual dependence upon MF3? Since these statistical and computational considerations do not preclude symmetric correlations between variables, the objection implies that the study provides just as good of evidence for the following *symmetric counterfactual:* Had ΔFC increased, then MF3 would have increased. Since explanation is widely thought to be asymmetric, mechanists may be tempted to deny that Adachi et al.'s model is explanatory.

We will argue that this objection is inconclusive given mechanists and autonomists' common ground. To see why, we underscore the importance of *background knowledge* in inferring explanations (Lipton 2004; Psillos 2007). For instance, even in well-designed experiments, background information is often required to infer the most plausible causal or mechanistic explanation. Similarly, we suggest that the following background knowledge helps to vindicate Adachi et al.'s inference:

> *AC's Explanatory Priority:* If *x* is statistically relevant to *y*, and *x* only represents AC, while *y* represents (at least some) FC, then it is *prima facie* more plausible to infer that *x* explains *y* than vice versa.

According to this principle, there is prima facie reason to believe that MF3 explains ΔFC rather than vice versa because MF3 only traffics in anatomical connectivity, while ΔFC traffics in both anatomical and functional connectivity. One motivation for this principle is that interventions on brain regions, neurons, and synapses are

[14] Additionally, causes and effects are frequently thought to be metaphysically distinct (Lewis 2000, 78), while topological explanantia and explananda frequently overlap, imply, and/or are identical to each other (Kostić 2019; Reutlinger 2018). Insofar as the same motifs figure in both MF3 and ΔFC, this would pose a further problem to construing this explanation etiologically.

[15] Perhaps situated mechanistic explanations are not etiological. However, then mechanists must show that situated mechanistic explanations do not face the problems we have raised with the organizational strategy in Section 4.2 and contextual mechanistic explanations in this section.

https://doi.org/10.1017/psa.2022.29 Published online by Cambridge University Press

easier to conceive of than interventions on voxels and synchronization likelihoods. Since interventions frequently serve as a guide to explanation, this gives AC networks their presumptive explanatory priority. That said, this priority is only prima facie; more rigorous testing and further theoretical considerations can overturn it. This background principle also accords with Adachi et al.'s reasoning. Based on their simulations and statistical analyses, they claim that MF3 "shapes" (Adachi et al. 2011, 1586, 1589–91) and "influences" (1586, 1591). ΔFC. They do not claim the converse, as the symmetric counterfactual suggests.

For mechanists who are sympathetic to AC's explanatory priority (Craver 2016; Povich 2015),[16] the objection is thereby defused: While Adachi et al.'s simulations and statistical tests are not sufficient unto themselves to license an explanatory claim, they become so in conjunction with this principle. Presumably, these mechanists take AC's explanatory priority to be a consequence of: (a) AC networks typically containing more mechanistic information than FC networks and (b) mechanistic information providing more reliable evidence for judging the plausibility of counterfactuals than information about functional connectivity. Our argument only requires AC's explanatory priority. So, these mechanists can only reject it on pain of also rejecting (a) or (b).[17]

But suppose that other mechanists would bite this bullet and reject AC's explanatory priority. We will argue that this entails that mechanistic and topological explanations are equally (in)vulnerable to symmetry problems. If asymmetry is a requirement on all correct explanations, then this means that both mechanists and autonomists are in trouble. On the other hand, if some correct explanations are symmetric, then autonomists can learn some valuable lessons from mechanists. For instance, Craver and Bechtel (2007, 553) claim that "all of the interesting cases of interlevel causation are symmetrical: components act as they do because of factors acting on mechanisms, and mechanisms act as they do because of the activities of their lower-level components." Indeed, Craver (2013) seems to leverage this point into taking constitutive and contextual mechanistic explanations to be simply two "perspectives" on the same system. So, seemingly harmless symmetries will exist when X constitutively explains Y and Y contextually explains X. Autonomists could devise analogues to constitutive and contextual mechanistic explanations to tolerate symmetries in a similar manner. On such a view, Adachi et al. would adopt one "perspective" in which MF3 explains ΔFC, but from another perspective, ΔFC explains MF3. So, regardless of whether AC enjoys explanatory priority over FC, autonomists are no worse off than mechanists with respect to symmetry.

In summary, while Adachi et al.'s explanation satisfies the Node and Edge Requirements, this does not make it a mechanistic explanation. The reasons for this are twofold. First, only topological properties are responsible for the explanandum; not mechanistic ones. Second, the explanation does not invoke the appropriate levels required for a constitutive mechanistic explanation. Thus, we have a counterexample to the most plausible mechanistic interpretation of topological explanations (MITE). Hence, absent some other MITE, we conclude that some topological explanations are non-mechanistic. Furthermore, this explanation not only resists characterization as a

[16] Section 6 discusses Craver's position on FC and AC in greater detail.

[17] Note that our reply to this objection commits autonomists neither to (a) nor to (b).

constitutive mechanistic explanation, but also as a *contextual* and as an *etiological* one. This suggests that other MITEs will face steep challenges going forward.

5. Autonomist account

A more convincing case for autonomism would provide a general account of topological explanation.[18] To that end, we take *a's* being *F* to topologically explain why *b is G* if and only if:

(T1) *a* is *F* (or approximately so);
(T2) *b* is *G* (or approximately so);
(T3) *F* is a topological property;
(T4) *G* is an empirical property; and
(T5) Had *a* been *F'* (rather than *F*), then *b* would have been *G'* (rather than G).

The account appears consonant with several others. It most strongly resembles Kostić's (2020) account, but omits certain details of his account that are not relevant to the tasks at hand. Proponents of more general counterfactual theories of noncausal explanation (e.g., those mentioned in note 22) should also be amenable to this account of topological explanation. We extend these views by using this account to unify topological explanations of both the mechanistic and non-mechanistic varieties. The former will satisfy the MITE and T1-T5; the latter will only satisfy T1-T5.[19]

Let us briefly motivate this analysis of topological explanation, and then show that it achieves the desired unification. The first two conditions, T1 and T2, are standard constraints on explanations—that the explanans and explanandum must be approximately true. Note that such a view still leaves ample room for idealization and other fruitful distortions of properties other than *F* and *G*. Furthermore, as T4 indicates, we assume that *b is G* is an empirical proposition, i.e., the sort of claim that can serve as a proper scientific explanandum.[20] Furthermore (and as our examples show), in many topological explanations "*a*" and "*b*" denote one and the same system.

The third condition, T3, distinguishes topological explanations from other kinds of explanations. Hence, it is crucial that we define a topological property. Let a *predicate* be topological if it correctly describes a graph (or subgraph) and occurs in some nontrivial theorem derived using only mathematical statements, including the characterization of the graph in terms of its vertices and edges. Then a graph's topological predicates denote its corresponding network's topological *properties.* Paradigmatically, topological properties concern quantifiable patterns of connectivity in a network. Clustering coefficient, average path length, and frequency of three-

[18] Importantly, with the exception of Kostić (2020), no autonomist provides anything like necessary and sufficient conditions for topological explanation. While meta-philosophical views may vary regarding the viability of such analyses, in the current context, such an approach evinces a more principled and rigorous unification of mechanistic and non-mechanistic topological explanations than other philosophical approaches to topological explanation. That being said, these other philosophical approaches appear well-suited for the different projects that our fellow autonomists have pursued.

[19] We postpone comparisons with other accounts of topological explanation for future work.

[20] T4 restricts topological explanations to empirical sciences; it excludes those in mathematics.

https://doi.org/10.1017/psa.2022.29 Published online by Cambridge University Press

node motifs are examples. As the examples above show, each of these topological properties is measurable and hence also empirical.

Finally, the fifth condition, T5, guarantees that the topological model is *explanatory*. As many others have noted, what distinguishes explanations from other kinds of representations is the former's capacity to support such change-relating counterfactuals,[21] or answer "what-if-things-had-been-different questions" (Jansson and Saatsi 2017; Reutlinger 2016; Woodward 2003, 2018).[22] Topological explanations also answer these questions, but they are distinctive in being underwritten by counterfactual differences in a system's topological properties. Such counterfactuals can describe what would happen if the system exhibited another topological property (in which case *F*' is contrary to *F*) or if it simply lacked its actual topological property (in which case *F*' is contradictory of *F*). Furthermore, we assume that only non-backtracking counterfactuals underwrite T5.

Crucially, T1-T5 allow some topological explanations to be mechanistic, but do not require all such explanations to work this way. For instance, both Watts and Strogatz's and Adachi et al.'s explanation readily fit our account. Begin with Watts and Strogatz. Small-worldness is a topological property that can be predicated of the nervous system of *C. elegans*. Thus, T1 and T3 are satisfied. Similarly, the extent to which information spreads throughout this nervous system is an empirical property that can be accurately predicated of this network. So, T2 and T4 are satisfied. Finally, in Section 2, we presented the counterfactual that would satisfy T5:

> Had the *C. elegans* neural network been regular or random (rather than small-world), then information transfer would have been less efficient (rather than its actual level of efficiency).

We saw that this explanation was both topological and, in virtue of satisfying the MITE, was also mechanistic. Turn now to Adachi et al.'s non-mechanistic topological explanation. Frequency of network motifs is a topological property that can be accurately predicated of macaque neocortices. In this context, ΔFC is characterized by observed correlations, specifically between different BOLD signal flows between different time series of brain areas, and by anatomical connections in length2-AC patterns which were confirmed in earlier studies (e.g., Honey et al. 2007). Hence, explananda will be true statements about these empirical facts. Thus, T1-T4 are satisfied. Finally, we have already rehearsed the relevant counterfactual in Section 4.1:

> Had macaque neocortices had a different frequency of three-node network motifs (rather than a different clustering coefficient, modularity, or frequency of two-node network motifs), then these neocortices' ΔFC would have been different (rather than its actual amount).

[21] Space being limited, we hope to address recent challenges to this "counterfactual assumption" (e.g., Khalifa et al. 2020; Lange 2019) in the specific context of topological explanations in future work.

[22] Jansson and Saatsi, Reutlinger, and Woodward discuss only one topological explanation: Euler's analysis of Königsberg bridges. We provide a more general account of topological explanation that covers both this simple example and more sophisticated topological explanations propounded by contemporary scientists. Also, as noted above, their concern is not with contrasting topological explanations with mechanistic explanations.

https://doi.org/10.1017/psa.2022.29 Published online by Cambridge University Press

So, Adachi et al.'s explanation also satisfies T5. However, unlike the explanation involving *C. elegans,* this explanation violated the MITE. Thus, we see that T1-T5 form the common core shared by both mechanistic and non-mechanistic topological explanations.

6. Functional connectivity

We have argued that some topological explanations are non-mechanistic. Moreover, we have also claimed that T1-T5 provide sufficient conditions for genuinely autonomous topological explanations, such as Adachi et al.'s. Craver provides a potential counterexample to this latter claim. To wit, he takes his counterexample to show that topological models are explanatory only insofar as they are mechanistic explanations. Hence, a more complete defense of topological explanations' autonomy should address Craver's challenge. To that end, we first present Craver's challenge, and then defend our core thesis that some topological explanations are not mechanistic explanations from this challenge.

Craver (2016, 704–6) argues that FC models are examples of topological models that are not explanations, chiefly because they do not represent mechanisms. As an illustration, we discuss Helling, Petkov, and Kalitzin's (2019) study of the relation between mean functional connectivity[23] (MFC) and the likelihood of an epileptic seizure (ictogenicity).

Craver observes that FC networks' nodes "need not ... stand for working parts," that is, for the entities that constitute a mechanism.[24] Rather, many FC models' nodes are conventionally determined spatiotemporal regions adopted mostly because they are "conveniently measurable units of brain tissue rather than known functional parts." For instance, Helling et al.'s FC model's nodes are readings from EEG channels, i.e., the electrodes measuring the brain's electrical activity.[25] EEG channels are spaced evenly—at increments of 10 or 20 percent of the distance from the bridge of the nose to the lowest point of the skull from the back of the head. This suggests that the spatial units represented by these nodes are merely conventional. As mentioned above, nontrivial conceptions of mechanism should distinguish entities and activities from spatiotemporal regions merely determined by convention. For instance, while pistons, gears, camshafts and the like are entities in the mechanism for a car's moving, each one-centimeter cube comprising a car is not an entity in that mechanism.

Similarly, while turning a crankshaft describes a piston's activity, whatever happens every two seconds to a piston does not (barring extraordinary coincidence of course.) Yet, the nodes in FC models pick out temporal units that are just as conventional as the spatial ones. In Helling et al.'s FC model, nodes are time series of readings from EEG channels. For each EEG channel, a time series was constructed by sampling its readings several times per second. So, in our parlance, it's quite clear that FC

[23] As mentioned above, functional connectivity's edges are correlations. Since correlations can be stronger or weaker, MFC is measuring the average strength of the correlations that exist between any two nodes in an FC network.

[24] In the same sentence, Craver makes the stronger claim that FC networks' nodes *do not* stand for working parts. However, as Adachi et al.'s explanation illustrates, this is not always the case.

[25] Craver focuses on FC models constructed from fMRI data, but his objections also apply to FC models constructed from EEG data.

https://doi.org/10.1017/psa.2022.29 Published online by Cambridge University Press

models violate the MITE's Node Requirement: They do not denote entities and activities constituting a mechanism.

Craver (2016, 705) also observes that FC models' "edges do not necessarily represent anatomical connections, causal connections, or communications," that is, they flout the MITE's Edge Requirement. Recall that this requires a topological explanation's edges to represent *interactions* between entities or activities. The edges in Helling et al.'s model are *synchronization likelihoods*, which are correlations between pattern recurrences in the time series data generated by two or more EEG channels. For mechanists, interactions cannot be mere correlations (Bechtel 2015; Craver and Tabery 2019; Glennan 1996).[26]

Craver takes these two points to plant the kiss of death for philosophers of explanation—guilt by Hempelian association:

> FC matrices are network models. They provide evidence about community structure in the brain. Community structure is relevant to brain function. But the matrices do not explain brain function. They don't model the right kinds of stuff: the nodes aren't working parts, and the edges are only correlations. As for the barometer and the storm, *A* is evidence for *B*, and *B* explains *C*, but *A* does not explain *C*.[27]

Craver is tapping on a powerful intuition: in most sciences, models consisting only of correlations are (at best) merely evidential but not explanatory of anything they represent. Call this the *barometer intuition*. In the case of FC models, the barometer intuition seems even more acute, for the correlations are between spatiotemporal units that lack causal roles in virtue of being merely conventional.

However, our view contradicts the barometer intuition. Helling et al. conducted prospective studies involving subjects with focal seizures either starting treatment with an anti-epileptic drug or undergoing drug tapering over several days. They collected EEG data from each patient in order to calculate each patient's MFC. Helling et al. found that MFC decreased for those who responded positively to their drug treatment, and increased for those who responded negatively.

Their model thereby satisfies T5. For instance, suppose that we ask why a patient responded negatively to anti-epileptic drugs. Then the relevant counterfactual would be:

> Had the patient's MFC decreased (rather than increased), then the patient would have responded positively (rather than negatively).[28]

Because MFC is a topological property, positive drug response is an empirical property, and the relevant statements are true, the model also satisfies T1-T4. Thus,

[26] Violating the Node and Edge Requirements entails violation of the Responsibility Requirement. Our discussion leaves this implicit throughout.

[27] In the original example, a storm's occurrence can be inferred from a barometer's reading. This is not an explanation, yet it satisfies Hempel's covering law model.

[28] More precisely, had the patient's MFC decreased (rather than increased), then the patient's ictogenicity would have been lower (rather than its actual value).

https://doi.org/10.1017/psa.2022.29 Published online by Cambridge University Press

Section 5's account claims that this is an explanation, which conflicts with the barometer intuition. Given how powerful the barometer intuition is, this appears to pose a serious problem to the account of topological explanation proposed in Section 5.

As we see it, the most straightforward autonomist reply embraces the barometer intuition and agrees with Craver about *FC models'* lack of explanatory power. However, it does not budge one inch on autonomism's more central point that *some topological explanations are not mechanistic explanations*. The argument is simple: Craver's point is that FC networks' topological properties are insufficient as *explanantia*. However, not all topological explanations work this way. For example, Adachi et al.'s topological explanation only features functional connectivity in the *explanandum*; other topological explanations do not use FC modeling at all. Hence, Craver's argument does not undermine Section 4's arguments.

Autonomists still must distinguish explanatory and non-explanatory topological models. This will require that they add further conditions to Section 5's unified account of topological explanation. Autonomists might contrast Adachi et al.'s explanation with Helling et al.'s FC model and propose that topological models are explanatory only if they satisfy T1-T5, plus:

T6. The topological model of *a* satisfies the Node and Edge Requirements.

Adachi et al.'s explanation satisfies T6, but (typical) FC models do not. Thus, we see that these views are autonomist in claiming that some of these explanations are *non-mechanistic* (namely those that violate either the Responsibility or Interlevel Requirement). However, we do not claim that this is the only way of supplementing T1-T5 or even of responding to Craver's challenge. Other approaches are possible and should be explored in future research.[29]

7. Conclusion

We began by noting that the debate between autonomists and mechanists suffered from imprecision. Specifically, the discussion lacked a clear account of how all topological explanations could be mechanistic. We have filled this gap, albeit in the service of providing examples of non-mechanistic topological explanations. We conclude that topological explanations sometimes swing freely of mechanistic considerations.

Of course, there is further work to be done. For instance, we have not discussed how graphs represent their respective networks, and this feeds naturally into the vibrant literature on scientific representation (Frigg and Nguyen 2017).[30] Whether similarity, structuralist, inferentialist, or some other account of representation best accords with topological explanations promises to be an interesting topic, broaching upon longstanding issues such as the applicability of mathematics.

Refining the kinds of counterfactuals that topological explanations ought to support is another exciting avenue of further development. As we see it, this has three crucial implications for advancing the position developed here. First, while we sketched an argument as to why some topological explanations are noncausal, future

[29] For instance, Kostić (forthcoming) argues, *pace* the barometer intuition, that FC models such as Helling et al.'s are indeed explanatory.

[30] See Hochstein (2016) for a relevant discussion of representation and (non-)mechanistic explanation.

https://doi.org/10.1017/psa.2022.29 Published online by Cambridge University Press

work should further investigate the link between topology and etiology. For instance, a prominent view is that causal explanations differ from noncausal ones in supporting counterfactuals involving interventions (Woodward 2003). Hence, a suggestive line of research is to explore the relationship between topological explanations and interventions.

Second, some have argued that topological explanations detached from any "ontic dependence relation" will fail to respect explanation's characteristic "directionality" (Craver 2016; Craver and Povich 2017). Whereas these ontic dependence relations used to be restricted to causal-mechanical relations, recent work has sought to broaden their range substantially (Povich 2018). Consequently, all topological explanations may still track with one of these broader ontic dependency relations. Alternatively, Kostić and Khalifa (2021) argue that nothing ontic is needed to account for topological explanations' directionality. Clarifying the precise nature of the counterfactuals involved in topological explanations helps to circumscribe what a more liberalized conception of ontic dependency relations entails, and thus proves useful in navigating these theoretical options (cf. Povich 2019).

Finally, further attention to the counterfactuals involved in topological explanations promises to address concerns that topological explanations are especially susceptible to classic Hempelian problems (such as asymmetry, irrelevance, and the like) because of their extensive appeal to mathematical derivations. We have already made a small contribution in assuaging this worry, by showing how our account can distinguish explanatory from evidential models, thereby blocking any tight analogy with Hempel's difficulties in preventing a barometer from "explaining" a storm. Moreover, using an analysis quite similar to our own, Kostić (2020) has outlined several ways that topological explanations are asymmetric. Nevertheless, assembling all of these points in a more systematic way would shed further light on topological explanations.

In closing, topological explanations are not merely a further chapter in the mechanist handbook. Attempts to incorporate all of them into a mechanistic framework are mistaken and fail to respect the unique features of these explanations. Moreover, doing so would foreclose several interesting questions to which philosophers of science would be well-served to attend.

Acknowledgements. We would like to thank Sara Green, Alexandros Goulas, Marc Lange, Arnon Levy, Charles Rathkopf, and two anonymous referees for their helpful discussions and comments on earlier drafts of this paper. We would also like to thank Tarja Knuuttila and her research group at the University of Vienna for their valuable feedback on this paper, which we presented in December of 2019.

Funding. Daniel Kostić would like to acknowledge funding by the Radboud Excellence Initiative. Kareem Khalifa would like to acknowledge funding from the American Council of Learned Society's Burkhardt Fellowship, "Explanation as Inferential Practice."

References

Adachi, Yusuke, Takahiro Osada, Olaf Sporns, Takamitsu Watanabe, Teppei Matsui, Kentaro Miyamoto, and Yasushi Miyashita. 2011. "Functional Connectivity between Anatomically Unconnected Areas Is Shaped by Collective Network-Level Effects in the Macaque Cortex." *Cerebral Cortex* 22 (7):1586–92. doi: 10.1093/cercor/bhr234.

Bechtel, William. 2009. "Looking down, around, and up: Mechanistic explanation in psychology." *Philosophical Psychology* 22 (5):543–64. doi: 10.1080/09515080903238948.

Bechtel, William. 2015. "Circadian Rhythms and Mood Disorders: Are the Phenomena and Mechanisms Causally Related?" *Frontiers in Psychiatry* 6:118. doi: 10.3389/fpsyt.2015.00118.

Bechtel, William. 2020. "Hierarchy and levels: analysing networks to study mechanisms in molecular biology." *Philosophical Transactions of the Royal Society B: Biological Sciences* 375:20190320. doi: 10.1098/rstb.2019.0320.

Bullmore, Ed, and Olaf Sporns. 2012. "The economy of brain network organization." *Nature Reviews Neuroscience* 13 (5):336–49. doi: 10.1038/nrn3214.

Craver, Carl F. 2001. "Role Functions, Mechanisms, and Hierarchy." *Philosophy of Science* 68 (1):53–74. doi: 10.1086/392866.

Craver, Carl F. 2007. *Explaining the brain: mechanisms and the mosaic unity of neuroscience*. Oxford: Clarendon Press.

Craver, Carl F. 2013. "Functions and Mechanisms: A Perspectivalist View." In *Functions: selection and mechanisms*, edited by Philippe Huneman, 133–58. Dordrecht: Springer Netherlands.

Craver, Carl F. 2016. "The Explanatory Power of Network Models." *Philosophy of Science* 83 (5):698–709. doi: 10.1086/687856.

Craver, Carl F., and William Bechtel. 2007. "Top-down Causation Without Top-down Causes." *Biology & Philosophy* 22 (4):547–63. doi: 10.1007/s10539-006-9028-8.

Craver, Carl F., and Mark Povich. 2017. "The directionality of distinctively mathematical explanations." *Studies in History and Philosophy of Science Part A* 63:31–38. https://doi.org/10.1016/j.shpsa.2017.04.005.

Craver, Carl F., and James Tabery. 2019. "Mechanisms in Science." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta. Stanford: Stanford University Press. https://plato.stanford.edu/archives/sum2019/entries/science-mechanisms/.

Darrason, Marie. 2018. "Mechanistic and topological explanations in medicine: the case of medical genetics and network medicine." *Synthese* 195 (1):147–73. doi: 10.1007/s11229-015-0983-y.

DiFrisco, James, and Johannes Jaeger. 2019. "Beyond networks: mechanism and process in evo-devo." *Biology & Philosophy* 34 (6):54. doi: 10.1007/s10539-019-9716-9.

Dupré, John. 2015. "Living Causes." *Aristotelian Society Supplementary Volume* 87 (1):19–37. doi: 10.1111/j.1467-8349.2013.00218.x.

Frigg, Roman, and James Nguyen. 2017. "Models and Representation." In *Springer Handbook of Model-Based Science*, edited by Lorenzo Magnani and Tommaso Bertolotti, 49–102. Cham, Switzerland: Springer International Publishing.

Glennan, Stuart. 1996. "Mechanisms and the Nature of Causation." *Erkenntnis* 44 (1):49–71.

Glennan, Stuart. 2017. *The new mechanical philosophy*. First edition. Oxford: Oxford University Press.

Green, Sara, Maria Şerban, Raphael Scholl, Nicholaos Jones, Ingo Brigandt, and William Bechtel. 2018. "Network analyses in systems biology: new strategies for dealing with biological complexity." *Synthese* 195 (4):1751–77.

Helling, Robert M., George H. Petkov, and Stiliyan N. Kalitzin. 2019. "Expert system for pharmacological epilepsy treatment prognosis and optimal medication dose prescription: computational model and clinical application." Proceedings of the 2nd International Conference on Applications of Intelligent Systems, https://doi.org/10.1145/3309772.3309775.

Hochstein, Eric. 2016. "One mechanism, many models: a distributed theory of mechanistic explanation." *Synthese* 193 (5):1387–407.

Honey, Christopher J., Rolf Kötter, Michael Breakspear, and Olaf Sporns. 2007. "Network structure of cerebral cortex shapes functional connectivity on multiple time scales." *Proceedings of the National Academy of Sciences* 104 (24):10240–5. doi: 10.1073/pnas.0701519104.

Huneman, Philippe. 2010. "Topological explanations and robustness in biological sciences." *Synthese* 177 (2):213–45.

Huneman, Philippe. 2018a. "Outlines of a theory of structural explanations." *Philosophical Studies* 175 (3):665–702. doi: 10.1007/s11098-017-0887-4.

Huneman, Philippe. 2018b. "Realizability and the varieties of explanation." *Studies in History and Philosophy of Science Part A* 68:37–50. https://doi.org/10.1016/j.shpsa.2018.01.004.

Illari, Phyllis McKay, and Jon Williamson. 2010. "Function and organization: comparing the mechanisms of protein synthesis and natural selection." *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 41 (3):279–91. https://doi.org/10.1016/j.shpsc.2010.07.001.

https://doi.org/10.1017/psa.2022.29 Published online by Cambridge University Press

Jansson, Lina, and Juha Saatsi. 2017. "Explanatory Abstractions." *The British Journal for the Philosophy of Science* 70 (3):817–44. doi: 10.1093/bjps/axx016.

Jones, Nicholaos. 2014. "Bowtie Structures, Pathway Diagrams, and Topological Explanation." *Erkenntnis* 79 (5):1135–55. doi: 10.1007/s10670-014-9598-9.

Khalifa, Kareem, Gabriel Doble, and Jared Millson. 2020. "Counterfactuals and Explanatory Pluralism." *British Journal for the Philosophy of Science* 71 (4):1439–1460. doi: 10.1093/bjps/axy048.

Kostić, Daniel. 2018. "The topological realization." *Synthese* 195 (1):79–98. doi: 10.1007/s11229-016-1248-0.

Kostić, Daniel. 2019. "Minimal Structure Explanations, Scientific Understanding and Explanatory Depth." *Perspectives on Science* 27 (1):48–67. doi: 10.1162/posc_a_00299.

Kostić, Daniel. 2020. "General theory of topological explanations and explanatory asymmetry." *Philosophical Transactions of the Royal Society B: Biological Sciences* 375:20190321. doi: 10.1098/rstb.2019.0321.

Kostić, Daniel. Forthcoming. "Topological explanations, an opinionated appraisal." In *Scientific Understanding and Representation: Mathematical Modeling in the Life and Physical Sciences*, edited by Kareem Khalifa, Insa Lawler, and Elay Shech. London: Routledge.

Kostić, Daniel, and Kareem Khalifa. 2021. "The directionality of topological explanations." *Synthese* 199 (5):14143–14165. doi: 10.1007/s11229-021-03414-y.

Kuorikoski, Jaakko, and Petri Ylikoski. 2013. "How Organization Explains." *EPSA11 Perspectives and Foundational Problems in Philosophy of Science*, Cham, Switzerland.

Lange, Marc. 2017. *Because without cause : non-causal explanation in science and mathematics*. New York: Oxford University Press.

Lange, Marc. 2019. "Asymmetry as a challenge to counterfactual accounts of non-causal explanation." *Synthese* 198 (4):3893–3918. doi: 10.1007/s11229-019-02317-3.

Latora, Vito, and Massimo Marchiori. 2001. "Efficient Behavior of Small-World Networks." *Physical Review Letters* 87 (19):198701. doi: 10.1103/PhysRevLett.87.198701.

Levy, Arnon, and William Bechtel. 2013. "Abstraction and the Organization of Mechanisms." *Philosophy of Science* 80 (2):241–61. doi: 10.1086/670300.

Lewis, David K. 2000. "Causation as Influence." *The Journal of Philosophy* 97 (4):182–97. doi: 10.2307/2678389.

Lipton, Peter. 1991/2004. *Inference to the best explanation*. Second edition. New York: Routledge.

Matthiessen, Dana. 2017. "Mechanistic Explanation in Systems Biology: Cellular Networks." *The British Journal for the Philosophy of Science* 68 (1):1–25. doi: 10.1093/bjps/axv011.

Povich, Mark. 2015. "Mechanisms and Model-Based Functional Magnetic Resonance Imaging." *Philosophy of Science* 82 (5):1035–46. doi: 10.1086/683438.

Povich, Mark. 2018. "Minimal Models and the Generalized Ontic Conception of Scientific Explanation." *The British Journal for the Philosophy of Science* 69 (1):117–37. doi: 10.1093/bjps/axw019.

Povich, Mark. 2019. "The Narrow Ontic Counterfactual Account of Distinctively Mathematical Explanation." *The British Journal for the Philosophy of Science* 72 (2):511–43. doi: 10.1093/bjps/axz008.

Psillos, Stathis. 2007. "The Fine Structure of Inference to the Best Explanation." *Philosophy and phenomenological research* 74 (2):441–48.

Rathkopf, Charles. 2018. "Network representation and complex systems." *Synthese* 195 (1):55–78. doi: 10.1007/s11229-015-0726-0.

Reutlinger, Alexander. 2016. "Is There A Monist Theory of Causal and Noncausal Explanations? The Counterfactual Theory of Scientific Explanation." *Philosophy of Science* 83 (5):733–45. doi: 10.1086/687859.

Reutlinger, Alexander. 2018. "Extending the counterfactual theory of explanation." In *Explanation beyond causation: philosophical perspectives on non-causal explanations*, edited by Alexander Reutlinger and Juha Saatsi, 74–95. Oxford: Oxford University Press.

Ross, Lauren N. 2020. "Distinguishing topological and causal explanation." *Synthese* 198 (10):9803–20. doi: 10.1007/s11229-020-02685-1.

Stam, C.J., B.F. Jones, G. Nolte, M. Breakspear, and P. Scheltens. 2006. "Small-World Networks and Functional Connectivity in Alzheimer's Disease." *Cerebral Cortex* 17 (1):92–99. doi: 10.1093/cercor/bhj127.

van den Heuvel, Martijn P., and Olaf Sporns. 2013. "Network hubs in the human brain." *Trends in Cognitive Sciences* 17 (12):683–96. doi: 10.1016/j.tics.2013.09.012.

Watts, Duncan J., and Steven H. Strogatz. 1998. “Collective dynamics of ‘small-world’ networks.” *Nature* 393 (6684):440–2. doi: 10.1038/30918.

Woodward, James. 2003. *Making things happen: a theory of causal explanation*. New York: Oxford University Press.

Woodward, James. 2013. “Mechanistic Explanation: Its Scope and Limits.” *Proceedings of the Aristotelian Society, Supplementary Volumes* 87:39–65.

Woodward, James. 2018. “Some Varieties of Non-Causal Explanation.” In *Explanation Beyond Causation: Philosophical Perspectives on Non-Causal Explanations*, edited by Alexander Reutlinger and Juha Saatsi, 117–140. Oxford: Oxford University Press.

Cite this article: Kostić, Daniel and Kareem Khalifa. 2023. “Decoupling Topological Explanations from Mechanisms.” *Philosophy of Science* 90 (2):245–268. https://doi.org/10.1017/psa.2022.29