Accepted for publication in *Synthese*.

How and when are topological explanations complete mechanistic explanations? The case of multilayer network models

Beate Krickel, Leon de Bruin, Linda Douw

Abstract. The relationship between topological explanation and mechanistic explanation is unclear. Most philosophers agree that at least some topological explanations are mechanistic explanations. The crucial question is how to make sense of this claim. Zednik (2019) argues that topological explanations are mechanistic if they (i) describe mechanism sketches that (ii) pick out organizational properties of mechanisms. While we agree with Zednik's conclusion, we critically discuss Zednik's account and show that it fails as a general account of how and when topological explanations are mechanistic. First, if topological explanations were just mechanism sketches, this implies that they could be enriched by replacing topological terms with mechanistic detail. This, however, conflicts how topological explanations are used in scientific practice. Second, Zednik's account fails to show how topological properties can be organizational properties of mechanisms that have a place in mechanistic explanation. The core issue is that Zednik's account ignores that topological properties often are *global* properties while mechanistic explanantia refer to *local* properties. We demonstrate how these problems can be solved by a recent account of mechanistic completeness (Craver and Kaplan 2020; Kohár and Krickel 2021) and use a multilayer network model of Alzheimer's Disease to illustrate this.

Keywords: network models, multilayer models, topological explanation, mechanistic explanation, mechanistic completeness, neuroscientific explanation

1. Introduction

In recent years, the idea that neuroscientists give *mechanistic explanations* has become increasingly popular. There is a broad consensus that mechanistic explanation involves structural and functional decomposition, i.e., breaking down a system into concrete parts and activities to identify the causal relationships that realize or constitute the phenomenon. However, it remains an open question to what extent mechanistic explanations can capture what is distinctive about research in systems neuroscience, in particular network analysis in systems neuroscience. Contrary to mechanistic explanation, network analysis seems to abstract away from concrete parts and activities to instead focus on the *topological* properties of connectivity patterns, with the aim of

explaining how they determine the behavioral dynamics of the systems exhibiting those patterns. Some authors have argued that network analysis differs from mechanistic explanation insofar as the explanatory power of these topological explanations is based on their topological properties and does not depend on concrete parts and activities in a specific organization (Huneman 2010; Rathkopf 2018; Kostić 2018). By contrast, authors such as Craver (2016) and Zednik (2019) have argued that topological explanations can only be genuinely explanatory if they can be understood as mechanistic explanations. However, it remains unclear how we can make sense of this. In this article, we will clarify under which conditions topological explanations are indeed mechanistic explanations by critically engaging with Zednik's account of topological explanations as mechanism sketches.

The article is structured as follows. In Section 2, we first offer a brief introduction to multilayer network analysis and discuss the idea of topological explanation. Also, we present an example of how multilayer network analysis can be used to explain cognitive decline in Alzheimer's Disease. In Section 3, we explain what mechanistic explanation entails. In Section 4 we give a critical analysis of Zednik's account of topological explanation as mechanism sketches that pick out organizational properties of mechanisms. We show that both aspects of this account are problematic because, first, it ignores the fact that topological properties are usually *global* properties while mechanistic explanantia refer to *local* organizational properties, and, second, topological explanations are not sketches of mechanisms that are made complete by replacing topological terms with mechanistic detail—they are good and complete mechanistic explanations the way they are. In Section 5, we put forward a recent account of mechanistic completeness by Kohár and Krickel (2021), according to which the completeness of a mechanistic explanation is measured relative to a contrastive explanandum. This account helps us to address the problems for Zednik's account by showing how topological properties as global organizational properties can be part of a complete mechanistic explanation. Thereby, we will show under which conditions topological explanations can be understood as mechanistic explanations. We will illustrate this by means of the Alzheimer's Disease example presented in Section 2.

2. Multilayer network models and topological explanation

2.1 Multilayer networks and multiplexes

A network is an abstract representation or model of a real-world system. The model consists of entities, called "nodes", which refer to specific properties of the original components of the system,

and the connection patterns between them, their links. Single models can be used to model only one type of connection between a predefined set of nodes (otherwise it cannot be handled by normal graph theory algorithms). To represent systems consisting of components with multiple types of connections, or other similar features, we need to model *layers* in addition to nodes and links. Different layers represent different aspects or features of the nodes that represent the components of the target system. In a multilayer model of commuter travel for instance, locations across a city may be the nodes. Consequently, we can partition the set of links into intralayer links, that is, links that connect nodes set in the same layer, such as the bus routes running between these locations in one layer, and the subway connections within another layer. Interlayer or coupling links are those links that connect nodes set in different layers. The resulting models are called "multilayer networks".

Multiplex networks are a sub-type of multilayer networks in which the same set of nodes is represented in every layer. The connections between nodes might be different in each layer.[1] In an *interconnected multiplex* network, nodes are connected only to themselves across all layers. The layer's connectivity, measured with respect to a specific definition of similarity (e.g., cross-correlation, spectral coherence, etc.) can be analyzed by focusing on activity in different frequency bands, time-varying activity and activity with respect to different tasks. Multilayer networks have been used to study various complex phenomena, appearing in various different types of systems such as social, biological, and transport systems (Mucha et al. 2010; Boccaletti et al. 2014; De Domenico et al. 2014). They are also increasingly used in network neuroscience to integrate different neuroimaging modalities sources of information on brain structure and function (e.g., structural and functional MRI; among other so-called "neuroimaging modalities"), or to study brain networks over different time points, among others (Vaiana and Muldoon 2020; De Domenico 2017).

2.2 Topological properties and topological explanation

In addition to just taking more data into account—whether to improve spatiotemporal resolution, provide a more comprehensive view, enhance measurement parameters or constrain certain data

[1] Note that multilayer/multiplex terminology describing particular systems may vary, as no universally conventional classification for the different variations of a network of networks exists. We adhere to the definitions put forward by Kivelä et al. (2014).

types—multilayer networks offer further opportunities to study the properties of connectivity patterns that shape the behavior of complex systems. That is, multilayer network analysis can explain the behavior of complex systems by referring to their *topology*. This type of explanation is called *topological explanation*.

A classic example might help to further illustrate what kind of properties topological explanations appeal to. In their seminal publication, Watts and Strogatz (1998) used network representations to explore the dynamics of infectious diseases. They used two central topological properties to assess the dynamics of this phenomenon. The first property is called the "characteristic path length". While the path length between any two nodes is the number of edges found on the shortest path between them, characteristic path length is a property of the whole graph, being the average path length over all pairwise combinations of nodes in the graph. The second property is the graph's "clustering coefficient", defined as the probability that two neighbors of a given node are themselves neighbors (where the term "neighbor" indicates the presence of a direct link between the nodes, regardless of spatial adjacency). On noticing that clustering coefficient and characteristic path length were very often anti-correlated in the empirical datasets they observed, with many networks having high clustering coefficients but very low characteristic path lengths, Watts and Strogatz called these networks *small-world networks.*

How does this shed light on the dynamics of infectious disease? In terms of their interactions, both social and spatial, human populations are highly clustered. Theoretically, high clustering could be expected to prevent a disease from reaching a large proportion of the total population quickly. However, Watts and Strogatz noticed that diseases can spread very quickly in a human population network because its characteristic path length is low. Since it only takes a small number of long-range connections to turn a highly clustered network with high path length into a small-world network with low path length, very small perturbations to the contact structure have an enormous impact on the dynamics of a disease. Indeed, Watts and Strogatz observed that a tiny change in the number of long-range disease transmission events makes diseases with low critical infectiousness rates capable of generating massive epidemics.

This example illustrates how topological properties can be used to explain the dynamics of a system. But what exactly are topological explanations? Topological explanation, according to Kostic (2020), describes how the mathematical properties of connectivity patterns in complex networks determine the dynamics of the systems exhibiting those patterns. Ross (2021) defines

topological explanations as "any explanation in which topology does the explanatory work" (Ross 2021, 9815). She argues that topological explanations are typically characterized as having three features (see Ross 2021, 9804–7): First, topological explanations appeal to the topology of the system (the relative position, organization, and structure of connections among entities in some domain) (Ross 2021, 9815), which captures a higher-level structure that abstracts away from various lower-level details. To say that the topology captures a "higher-level" structure means that this structure can be instantiated or realized by a variety of different physical or microstructural details. For example, many different systems can be described as small-world networks, such as websites, electric power grids, and airport networks. Second, the topology of the system is usually taken to capture a structure that is non-causal in the sense that it lacks (temporal) information that causal structures necessarily contain (ibid.). Small-worldness, to continue the example, lacks not only information about the causal process of how, for example, diseases spread, but it also lacks information about the nature of the things that make up the system's structure (i.e., the individuals and their causal interactions). A third feature of topological explanation concerns the dependency relations that specify how the explanandum is dependent on the explanans (Ross 2021, 9806). Whereas, in causal explanation, the dependency relation is typically identified and verified empirically, in the case of topological explanation the dependency relation is provided by mathematical derivation (Ross 2021, 9815–16). Thus, as Ross points out, once we understand the topology of, for example, a society, we can apply mathematical understanding to answer the question why diseases spread and turn into epidemics the way they do.

Ross stresses that, although most topological explanations exemplify these three features, there are topological explanations that diverge from this standard picture—which she calls "causal–topological explanations." These explanations are based on network models in which the edges either directly represent causal relationships between these entities or at least contain causal information (Ross 2021, 9808–10). In these explanations, the dependency relation between the explanans and the explanandum "involves a significant amount of empirical information" (ibid.), and, therefore, cannot be mathematically derived from the explanans (Ross 2021, 9817).

In what follows, we will present an example of such a causal topological explanation that further illustrates Ross's ideas and that will serve as the test case for our mechanistic analysis in Section 5.

2.3 An explanation of Alzheimer's Disease

Our example is an explanation of cognitive decline in patients with Alzheimer's Disease (AD). In addition to protein misfolding, AD is characterized by several neurophysiological changes including slowing of the alpha rhythm and suboptimal connectivity and network organization.[2] In particular, patients with AD show dysfunction of regions in the brain that are normally densely connected. These regions—so-called "hubs"—are thought to be responsible for the overall functioning of the entire brain. Whether a node is a hub depends on its *centrality*, i.e., its number of connections relative to the number of connections of the other nodes in the network.[3] Centrality is a gradual topological property rather than a binary one: there are no strict cut-offs for hubs versus non-hubs. Computational modeling has shown that intervening into central nodes (either by removing or altering them) has a bigger effect on overall network integrity than intervening into more peripheral nodes (Alstott et al. 2009; Honey and Sporns 2008). This suggests that the extent to which node dysfunction happens in central regions may be relevant to the level of cognitive decline in AD.

Several theoretical frameworks have been proposed to specify how diseases like AD may unfold across the network, and how specific network topological changes may explain cognitive decline. A very promising framework is the "cascading network failure" hypothesis (Stam 2014), which postulates that more central regions of the brain are particularly vulnerable to neurodegeneration and other brain pathology due to their high premorbid metabolic cost and their role as a relay station within the entire brain network. The idea is that, initially, local brain dysfunction (caused, for instance, by the cellular death associated with adjacent misfolding proteins being present) does not propagate through the network due to the integrative and adaptive role of more central regions. These central regions/hubs are accustomed to varying demands, as they connect to a variety of other brain regions and thus receive dynamic input anyways. This phase may clinically be characterized by minimal or mild cognitive impairment, since only local

[2] The cause of AD was hypothesized to be a combination of the misfolding of proteins tau and amyloid beta proteins within the brain that develops with a typical spreading pattern. However, the timescale over which cognitive decline develops and the level of impairment to be expected at any given moment is heterogeneous. In other words, some AD patients have cognitive deficits at a certain time point during the course of their disease, and others at the same time point do not. This heterogeneity cannot be explained by protein-level pathophysiology.

[3] Several graph metrics can be used to measure centrality. The centrality mentioned is degree centrality, which quantifies the total number of connections of a node relative to the total number of connections of other the nodes in the network. Other centralities include eigenvector centrality, which not only takes the number of connections that a node itself has into account, but also incorporates the number of connections of its neighbors.

brain dysfunction occurs in the network. However, as central regions get more and more burdened by increasing cellular pathology, "hub overload" is thought to occur. That is, local increases in activity and connectivity anywhere in the network result in a larger burden on those most connected regions in the network, because the increase in information processing load rapidly spreads throughout the network via these well-connected integrators. This phase of overload is hypothesized to mark the point where Alzheimer's dementia is diagnosed (Jones et al. 2016). Of particular interest is that this overload happens in a phase-transitional manner, instead of being a slow or gradual process (Watts 2002). This is exactly what we see in AD patients. What is more, such abrupt transitions are seemingly more abundant in multilayer networks than in single-layer networks according to mathematical studies (Boccaletti et al. 2014; Baxter et al. 2012).

A recent study by Yu and colleagues (2017) demonstrated how a multilayer network approach can be used to explain AD symptoms. They implemented multilayer network models based on often-investigated frequency bands in magnetoencephalography (MEG), ranging from slow delta (0.5–4 Hz) to fast gamma (50–80 Hz) brain oscillations, and subsequently performed single-layer and multilayer network analyses. For their definition of nodes, Yu and colleagues used a widely accepted atlas with a parcellation of brain regions spanning the entire gray matter and subcortex of the brain, consisting of 90 nodes. Using this atlas ensured that the same anatomical brain regions were used as nodes across all subjects and layers. The different layers of the multilayer model corresponded to the five frequency bands (delta, theta, lower and upper alpha, beta). The intralayer connections between these nodes were defined by phase-based synchronization (with weights ranging between 0 and 1, where "0" corresponds to "no synchronization" and "1" to "full/complete/perfect synchronization") between oscillations measured in the frequency bands. For these intralayer connections, it was assumed that synchronized oscillatory behavior at the level of brain regions reflects (propensity for) functional communication between them (Yu et al. 2017, 1469). To construct multilayer network models for each subject (AD patient and control), interlayer connections were created between the same nodes/brain regions across layers (resulting in an *interconnected multiplex*, as explained in the previous section) with a weight of 1, so no oscillatory or synchronization information was used to

weight these connections.[4] See Figure 1 for a (highly simplified) illustration of the resulting models.

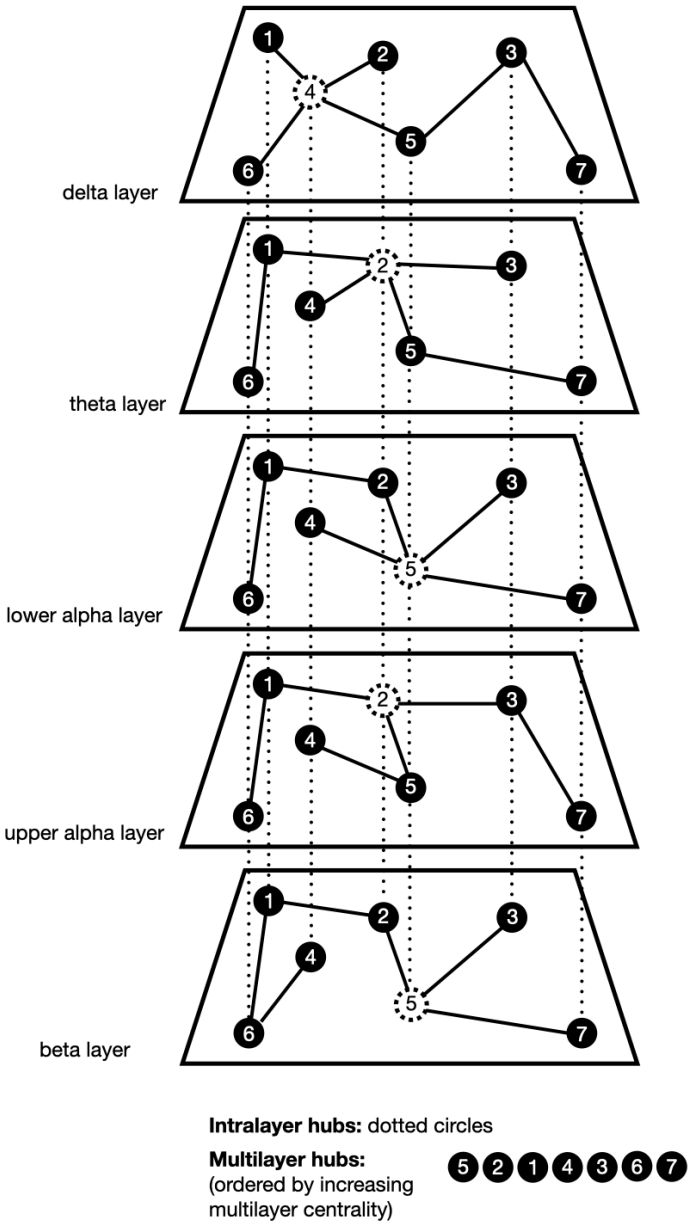

Figure 1 An illustration of the multilayer networks (interconnected multiplexes) created by Yu and colleagues for each subject. Layers represent different frequency bands. The nodes (circles) represent the 90 brain parcels/regions that Yu and colleagues focused on. Straight lines between nodes at the same layer represent statistical correlation between the activity patterns of the connected regions. Dotted lines indicate interlayer connections that were created between the same nodes across layers with a weight of 1. *Intralayer hubs* (dotted circles) are those nodes that are most highly connected within one layer compared to the other nodes at this layer. *Multilayer hubs* (see bottom of Figure 1) are those nodes that are most highly connected compared to the other nodes taking all layers into account. As multilayer centrality is a gradual property, nodes 1 to 7 can be ranked according to their multilayer centrality.

[4] Although it is technically possible to incorporate cross-frequency synchronization instead (see Tewarie et al. (2016)), this is computationally expensive, and it remains unclear what the biological meaning of this type of synchronization between frequency bands is.

The comparison between patients and controls was based on assessing centrality in these individual multilayer network models, by calculating *multilayer centrality*. Instead of taking only the number of connections within a layer (frequency band) into account, multilayer centrality also attends to the aggregated connectivity of a node across all layers (see Figure 1). Finally, a disruption score was calculated by subtracting the average multilayer centrality of the healthy controls for each specific region from this individual's multilayer centrality value of that region. This value thus indicates the *network disruption value*, i.e., how disrupted the level of centrality of each region is in each subject's brain network in comparison to the default healthy situation. Using this approach, Yu et al. found that

1. network disruption values were higher in AD patients than in healthy controls. This difference was particularly significant when considering multilayer disruption, as opposed to single-layer disruption values. Moreover, there was a correlation between a region's centrality in healthy controls, and the degree of hub disruption in AD patients, such that the regions that are normally most central in the healthy multilayer network were more disrupted in patients with AD. Again, this correlation was particularly significant when using multilayer disruption values, as opposed to single-layer disruption values;
2. multilayer centrality was correlated with cognitive function in AD patients, such that lower multilayer centrality values of regions that usually have higher centrality (i.e., a selection of regions based on the healthy controls) go hand-in-hand with poorer cognitive scores from patients.

These results underline the value of the multilayer network approach in investigating how cognition breaks down in AD. Although correlational at this point, the study offers a valuable description of the complete neurophysiological network model in AD and may explain why results from previous single-layer network studies were inconsistent. Moreover, there is a vast literature on the impact of damage on multilayer networks: this largely mathematical literature offers analytical evidence towards explaining the behavior of (subparts of) the system once damage occurs in particular regions. The study by Yu et al. should therefore be complemented with a modeling study making use of such analytical insights in combination with a longitudinal empirical study of the progress of AD in order to assess whether theory on failures in multilayer networks indeed explain AD progression of cognitive deficits in AD.

This example illustrates the potential power of topological explanation in clinical neuroscience. Yu and colleagues explain cognitive decline in AD (the explanandum) in terms of a decreasing multilayer centrality for brain regions in AD patients that were most central in healthy subjects (the explanans). According to Ross (2021), this is a topological explanation as the explanans refers to a topological property of the brains of AD patients, *decrease of multilayer centrality*. Multilayer centrality is a topological property, i.e., a network property that is characterized mathematically and that is multiply realizable.

3. Mechanistic explanation

As mentioned in the introduction, some authors have argued that topological explanations are distinct from mechanistic explanations insofar as their explanatory power is solely based on their topological properties rather than mechanistic details (Huneman 2010; Rathkopf 2018; Kostić 2018). Zednik (2019), however, has pointed out that the assumption that there is a strict distinction between topological and mechanistic explanations is misleading, and that some topological explanations can in fact be understood as mechanistic explanations. Before we are in the position to assess whether and how topological explanations such as the AD study do indeed qualify as mechanistic explanations, we first need to spell out the mechanistic explanatory approach in more detail.

To mechanistically explain a phenomenon, roughly means to describe the mechanism that is responsible for the phenomenon. This is the core claim of the so-called *new mechanistic account*. The details of this account concern the analysis of the different components of the core claim. They vary from author to author and are still a matter of lively discussion. Thus, what we present here is, to some degree, an opinionated summary of the existing literature. The key questions are:

(i) What is a *mechanism*?
(ii) What is a *phenomenon*?
(iii) What is meant by "*responsible*"?

With regard to the first question, a more-or-less standard characterization has emerged in the mechanistic literature: Mechanisms are entities and activities in a specific organization (Craver 2007b; Illari and Williamson 2012; Glennan 2017). The entities and activities are those that are commonly accepted by contemporary biology, chemistry, and physics, such as ions, cells,

molecules, organs, and organisms, diffusion, movement, attraction, repulsion, collision, binding, and transmission. The organization concerns the *spatial* organization of entities (Craver 2007b, 137–38): entities have a certain size, they are located at specific locations, they have a certain orientation, are in contact with other entities, and are connected to others. It also concerns the *temporal* organization of the activities (Craver 2007b, 138): activities occur in a certain order, they have a certain duration and rate. Finally, the organizational aspect concerns the *active* organization of the entities and activities (Craver 2007b, 136–37): entities act in certain ways and not in others, they causally interact with certain entities but not with others, they act in cooperation or in competition, and they do so in specific ways.

Whether a given entity or activity is a component of a mechanism depends on whether the entity or activity is relevant for the phenomenon for which the mechanism is supposed to be responsible. The phenomenon is usually taken to be a property or behavior of a system (Craver 2007b; Krickel 2018a). Examples are the firing of a neuron, the stretching of a muscle, the navigating of a rat.

The responsibility relation between a mechanism and a phenomenon is causation (you can plug in your favorite account of causation) or constitution, where the former is taken to involve wholly distinct relata, while the latter holds between wholes and their parts. What exactly mechanistic constitution is, is still a matter of debate. The most prominent account is Craver's *mutual manipulability account* (MM) (Craver 2007b; 2007a):

> MM: X's ϕ-ing (a mechanism's component) is constitutively relevant for S's ψ-ing (a phenomenon) if and only if[5]:
>
> a. X's ϕ-ing is a spatiotemporal part of S's ψ-ing, and
> b. one can change S's ψ-ing by changing X's ϕ-ing, and one can change X's ϕ-ing by changing S's ψ-ing.

There is an on-going debate on how to spell out the details of this account (Krickel 2018b; Romero 2015; Kästner 2017; Baumgartner and Gebharter 2016; Baumgartner and Casini 2017; Baumgartner, Casini, and Krickel 2020; Harinen 2018; Prychitko 2021; Craver, Glennan, and

[5] It is unclear whether the conditions are only sufficient conditions or also necessary conditions (see, e.g., Craver, Glennan, Povich (Craver, Glennan, and Povich 2021, n. 7).

Povich 2021). We do not want to go into the details of this discussion here. What is important for our purpose is that mechanistic constitution minimally implies that the mechanism is *local* to the phenomenon (Illari and Williamson 2011): the mechanism's components are spatiotemporal parts of the phenomenon, and there is some kind of mutual dependence between the mechanism's components and the phenomenon. This mutual dependence is usually established with help of top-down and bottom-up interventions. The bottom-up intervention is an intervention into a component with respect to the phenomenon; the top-down intervention is an intervention into the phenomenon with respect to a component. Interventions into phenomena, thereby, are usually conceived of as interventions into inputs of mechanisms (Baetu 2012; Harinen 2018; Craver, Glennan, and Povich 2021).

Given these three notions, there are two different general conceptions of mechanistic explanation: the ontic and the epistemic conception.[6] According to some defenders of the new mechanistic approach, explanation is primarily an *ontic* matter: it is the mechanism itself that does the explaining. Another view is that explanation is an *epistemic* matter in the sense that explanations are representations or descriptions of things in the world, i.e., the mechanisms and the phenomenon. In this view, not just any representation or description of the mechanism (i.e., of the mechanistic components that are constitutively relevant for the phenomenon) is taken to be explanatory. Rather, further pragmatic factors, such as interests and goals, constrain which descriptions of mechanisms are explanatory for a given phenomenon and which are not.

The ontic account is often identified with the characterization given above: a mechanism (a mind-independent thing in the world) explains a phenomenon (a mind-independent thing in the world) because the former causes or constitutes the latter (a mind-independent relation). The epistemic account (or its different versions) relies on the same characterization of mechanism, phenomenon, and constitution/causation. However, it denies that this is sufficient for genuine scientific explanation. To explain a phenomenon, one has to find the right sort of description of the relevant mechanism. How this is to be spelled out in detail is a matter of dispute; we will come back to it in Section 5.

[6] Note that the ontic and the epistemic view are compatible with one another: "to explain" is simply ambiguous in at least these two senses (Craver 2014).

4. Topological explanations as mechanistic explanations: a critical assessment of Zednik's account

Zednik (2019) argues that topological explanations are mechanistic explanations if they are mechanism sketches that highlight organizational properties of mechanisms. Zednik's argument is a reply to Craver's criticism of a certain type of topological explanation, i.e., explanations based on functional network models (Craver 2016). According to Craver, such explanations are non-mechanistic and non-explanatory because they do not pick out working parts and fail to describe real causal connections. For example, a functional network model of neuroscientific phenomena measured with functional magnetic resonance imaging (fMRI) is based on more-or-less arbitrary parcellations of the brain and relies on statistical information about the interdependencies among the regions' activation patterns rather than causal information. However, Zednik argues that this is compatible with topological explanations being mechanism sketches (Zednik 2019, 15). Zednik's main point is that the limitations put forward by Craver are due to practical constraints rather than explanation-inherent. Furthermore, while statistical information does not yet prove the existence of causal connections, it can convey information about the causal organization of the underlying mechanism (Zednik 2019, 15–16). Similarly, while not mapping exactly on working parts, pragmatic parcellations do approximate working parts (Zednik 2019, 16–18).

As already mentioned above, Zednik's account formulates two conditions for when a topological explanation is a mechanistic explanation: first, the topological explanation is a *mechanism sketch* that, second, refers to *organizational properties of the mechanism*. We will critically discuss both of these conditions, beginning with the second one.

As mentioned in the previous section and as highlighted by Zednik, organization is a crucial aspect of mechanisms as the same entities and activities might bring about completely different phenomena if they are organized differently. Zednik observes that despite the centrality of organization in the notion of a mechanism, the issue of organization has not been the focus of the new mechanistic thinking. Most importantly, Craver's original mutual manipulability account of constitutive relevance (see Section 4) only applies to components/acting entities (X's ϕ-ings) and does not mention organizational properties. To solve this problem, Zednik extends Craver's account:

> **MM-O:** A topological feature is an organizational property of a mechanism [for S's ψ-ing] if one can change the behavior of the mechanism as a whole [i.e., S's ψ-ing] by intervening to change that topological feature, and one can change the topological feature by intervening to change the behavior of the mechanism as a whole [i.e., S's ψ-ing]. (Zednik 2019, 22)

As MM is an account of mechanistic constitution, MM-O, if successful, shows that topological properties can be constitutively relevant for a system's behavior (S's ψ-ing) by being organizational properties of the mechanism for S's ψ-ing. Thus, according to Zednik, topological properties are potential organizational properties of mechanisms—and they are actual organizational properties of mechanisms if they satisfy MM-O, i.e., if changing the phenomenon (S's ψ-ing) will change the topological property and changing the topological property will change the phenomenon.

MM-O, however, is problematic. MM-O indeed only mentions the second condition of the original mutual manipulability account (see previous section). An analogue to the first condition—"X's ϕ-ing is a spatiotemporal part of S's ψ-ing"—is not mentioned. As a consequence, MM-O identifies topological features as properties of a mechanism that are not necessarily properties of the relevant mechanism. To illustrate this, consider the following example: assume that there are two neural networks A and B, each with a certain topology. The networks are connected such that, depending on their oscillation patterns, they can activate each other. Whenever network A activates network B, this will change B's topology and whenever B activates A, this will change A's topology. Now, assume we want to explain some behavior of A, say, the fact that it oscillates with a certain frequency. Given the setup of the two networks, it will be true that there is a way of changing B's topology (changing it such that oscillations of a certain type occur) by which the frequency of A's oscillations will be changed; and there is a way of changing A's oscillation frequency (changing it such that A activates B) that will change B's topology. According to MM-O, this suffices to show that B's topology is constitutive for A's oscillating with a certain frequency. However, while changing B's topology may be a *cause* of A's oscillating with a certain frequency, it is not *constitutive* for it, i.e., it is not part of the relevant mechanism. Only A's topology is constitutively relevant for A's oscillating with a certain frequency. Thus, we need to make sure that topological properties satisfy the first condition of the mutual manipulability account as well.

As properties do not have any specific locations (unless you think of them as tropes), it makes most sense to assume that the location of a property is the location of the entities that instantiate the property (see Kohár 2022). Based on these considerations, MM-O should be revised as follows:

> MM-O*: A topological property T is an organizational property of a mechanism for S's ψ-ing if the (acting) entities that instantiate T are spatiotemporal parts of S's ψ-ing, and one can change S's ψ-ing by intervening to change T, and one can change T by intervening to change S's ψ-ing.

This revision of MM-O, however, is problematic as it is impossible for topological properties to satisfy it. The reason will become clear from our consideration of the second problem.

The second problem for Zednik's idea that topological explanations are mechanistic because they refer to organizational properties of mechanisms can be illustrated with the help of the distinction between *global* and *local* organizational properties. Global organizational properties are properties of a system as a whole, whereas local organizational properties are properties of particular parts of a system[7]. As explained in the previous section, the organizational properties that characterize a mechanism are properties of the entities and activities that make up the mechanisms. Thus, they are local (not global) organizational properties in the sense just specified. Surely, many mechanists implicitly or explicitly accept that mechanisms as wholes have global organizational properties as well (e.g., mechanisms can be feedback loops). To date, however, it remains unclear how these types of organizational properties relate and how to integrate global organizational properties into the mechanistic account.

There are at least three interrelated challenges that show why this clarification is urgent and non-trivial:

a) The divide between global properties and local properties matches that between the explanans and explanandum. According to the mechanistic account, it is the interactions between the *mechanism's components and their local properties* that do the explaining and it is the *properties and behaviors of the mechanisms as a whole*

[7] Note that the term "local" here has a different meaning than how it is used in the Section 3. In Section 3, mechanisms and their components are said to be *local* to the phenomenon (following Illari and Williamson 2011). This expresses the idea that all the mechanism's components are *spatiotemporal parts of the phenomenon*. The contrast would be "outside", or "external", not "global."

that is to be explained (Craver 2007; Craver, Glennan, and Povich 2021; Kaiser and Krickel 2017).[8] Therefore, prima facie, global organizational properties cannot be part of the explanans but only of the explanandum of a mechanistic explanation.

b) Local organizational properties satisfy MM-O* and can therefore be parts of mechanistic explanantia. MM-O*, thus, nicely captures the original idea of the new mechanists. However, qua global properties, global organizational properties cannot satisfy the first condition of the mutual manipulability account as they are not properties of components of mechanisms but of mechanisms as wholes. Thus, they cannot satisfy MM-O*. Dropping the first condition of the mutual manipulability account is not a solution, as we show above.

c) Global organizational properties are multiply realizable by and thus not reducible to local organizational properties. This blocks an easy solution to problem (b) (see below).

The challenges arising from the local/global distinction are general concerns for the new mechanists. For Zednik's account, the challenges are especially pressing as topological properties are, if at all, *global* properties of mechanisms. Thus, one needs to show how topological properties can be parts of the explanans of a mechanistic explanation even though they are global properties of mechanisms and not local properties of the mechanism's components.

Zednik might reply that topological properties are nothing but local organizational properties combined in the right way, i.e., that global organizational properties can be reduced to local ones. For example, one might argue that the property of small-worldness can be reduced to the relative spatial relations and causal connections between specific entities. This, however, cannot be true. While all healthy brains share topological features, their mechanistic organization could be quite different—they may be composed of different numbers of neurons, their neurons may be spatially arranged in different (relative) locations, they have different orientation, they have different connection densities, and so on. In other words, global topological organizational properties are

[8] The expression that in constitutive mechanistic explanations what is explained is the "mechanism as a whole" is a common assumption in the mechanistic literature (Craver 2007; Glennan 2017; Craver, Glennan, and Povich 2021). Kaiser and Krickel (2017) and Krickel (2018) have argued that this assumption is problematic. For reasons of space, we do not want to go into detail here. For argument (a) to work, we need not commit to the idea that all explananda of mechanistic (constitutive) explanations concern properties or behaviors of mechanisms as wholes. All we need is that if a property is a property of the mechanism as a whole, it cannot be part of the mechanistic-constitutive explanation of a property or behavior of the same mechanism. It can, however, be part of the explanandum.

multiply realizable by different combinations of local organizational properties. Hence, the former cannot be reduced to the latter.

Another way to solve the problems arising from the globalness of topological properties might be to refer to the first condition of Zednik's account: topological explanations provide *mechanism sketches*. Indeed, topological explanations are, while explanatory, still incomplete. Once the sketchy, topological aspects of the explanation are replaced by or reduced to local organizational properties, one has a complete mechanistic explanation. In this picture, terms referring to topological properties would be *filler terms*. An example of a filler term is the term "channel" in the explanation of the action potential: "[a]t the time the idea of a channel was viewed with skepticism. It was merely a filler-term for an activity or mechanism to be named later" (Craver 2007b, 58). That is, filler terms are terms that are used "as place-holders for future work" and, at some point, they have to be replaced "with some stock-in-trade property, entity, activity, or mechanism" (ibid., 113). In that sense, topological terms would have to be replaced by more detailed descriptions of the local organizational features of mechanisms.

If topological terms were filler terms, this would imply that, for any true topological explanation, there would be an explanation in purely local organizational terms, and this explanation would be better because it includes the mechanistic detail in place of the topological filler terms. However, this is not in line with what we see in actual scientific explanatory practice. The explanation of AD does not mention any particular connections between nodes. It mentions terms like "decrease of multilayer centrality", and it is not clear that this explanation would be improved by adding detail about specific connections. For example, if one were to describe the exact number of nodes, their relative positions, and connections for each multilayer network for each patient, this would not add anything relevant for the explanation of cognitive decline in AD—what is relevant is that all these multilayer networks show a decrease of multilayer centrality. Of course, this is not to say that the explanation in terms of a decrease of multilayer centrality is already complete—there may be further details that are explanatory relevant for the given explanandum (for example, the exact degree of decrease of multilayer centrality may be relevant and whether there are any differences in the degree among AD patients). Nevertheless, treating "decrease of multilayer centrality" as a filler term and replacing it by a description that lists all the mechanistic details would not be any more informative or explanatory than the explanation that uses topological terminology. Additionally, due to multiple realization, it would be unclear which

local organizational properties are to be filled in from the set of possible realizers. This shows that topological terms such as "decrease of multilayer centrality" are not just filler terms.

Thus, the question remains as to how and when topological explanations can be mechanistic explanations. To show that, one needs to show how global topological properties can play a role in mechanistic explanations despite their globalness. Also, one must make sense of the idea that topological terms are not just filler terms awaiting replacement by local organizational terms but that topological explanations are already complete. In the next section, we will present a recent account of mechanistic completeness, based on which we will show that topological explanations can be complete mechanistic explanations.

5. Filling the gaps: Mechanistic completeness

5.1 Mechanistic completeness

One crucial issue for the new mechanistic account of scientific explanation is to specify when a mechanistic explanation is complete. Such a specification is necessary to avoid the implausible consequence that, say, the explanation of spatial memory has to go down to the most fundamental physical level. Furthermore, if one wants to save the idea that mechanistic explanations are objective or ontic, this "bottoming-out" of mechanistic explanation should not be simply a matter of taste or interest but should be based on objective criteria. One recent account of mechanistic explanation has been put forward by Kohár and Krickel (2021), who base their ideas on Craver and Kaplan's (2020) idea to evaluate mechanistic completeness relative to a contrastive explanandum. Here, we want to use Kohár and Krickel's account to show that some topological explanations are indeed complete mechanistic explanations.

Mechanistic completeness, according to Craver and Kaplan (2020) has to be evaluated relative to a contrastive explanandum. While this is a deviation from the original account (see Section 3), this idea is not foreign to the new mechanistic thinking (Craver and Kaplan 2020[9];

[9] The crucial point of Craver and Kaplan (2020) is that to determine the completeness of a mechanistic explanation, one has to evaluate a given explanation based on to what degree it answers a specific contrastive explanandum. However, Craver and Kaplan still adhere to the ontic conception of mechanistic explanation. Kohár and Krickel (2021) show that contrastive explananda require an epistemic reading of mechanistic explanation: contrastive explananda are not things in the world but representations thereof. Also, contrastive explananda come along with contrastive explanantia - which in the same way are not things in the world but representations thereof. Furthermore, Kohár and Krickel argue that whether what they call "vertical completeness" of mechanistic explanation is reached is to be determined based on whether one has found "crucial points of intervention" i.e., point where an intervention is "such that the intended phenomenon is produced in the most economic fashion (with minimal effort)" (Kohár and Krickel 2021, 421). Although Kohár and Krickel do not go into detail about what is meant by "most economic fashion", it is

Woodward 2011). One way to specify this idea is that the explanandum takes the form of a contrastive question "Why is G the case rather than F?" (or: "Why does variable *X* have value *p* rather than *q*?") that picks out a specific feature of the phenomenon (i.e., G). The explanans is a description of the differences between the mechanism that causes or constitutes the actual phenomenon and the mechanism that causes or constitutes the contrast phenomenon that would be/have F. Thus, the explanans picks out *only certain aspects* of the actual mechanism.

Kohár and Krickel (2021) develop an explicit account of mechanistic completeness based on the ideas put forward by Craver and Kaplan (2020). They provide a detailed analysis of how contrastive mechanistic explanation in the sense characterized above proceeds. Their account can be illustrated with the help of the following example. If you want to know why the action potential of a specific (type of) neuron always peaks at a value of, say, 30 mV rather than some higher value, you must do the following (note that this procedure is a regulative ideal—in practice, the procedure is much messier and must involve inductive reasoning):

1. Identify the mechanism that constitutes the action potential of that (type of) neuron: find all entities and activities and their organization that are constituents of the firing of that neuron.
2. Identify the mechanisms that constitute action potentials that peak above 30 mV.
3. Identify the mechanism in step 2 that is maximally similar to the mechanism in (1).
4. List all the differences between the mechanism in (1) and the mechanism in (3) that are shared by all mechanisms in (2) (see Kohár and Krickel (2021) for details).

The list resulting from step 4 will be the explanans, i.e., the answer to the question, "Why does the action potential of this neuron always peak at 30 mV rather than a higher value?" In other words, it identifies the difference-maker(s) for peaking at 30 mV versus peaking at a higher value. The addition "that are shared by all mechanisms in (2)" in step 4 is important for the following reason: the contrast phenomenon can have many different mechanisms. In the given example, the contrast phenomenon may be due to mechanisms that produce peaks at 31 mV, 50 mV, 70 mV, 3967 mV, and so on. In order to answer why the actual value is 30 mV rather than something bigger, we are

plausible to assume that this may well depend on pragmatic factors that minimize the effort (such as the availability of the right kind of experimental tool).

not interested in the differences between the actual mechanism and, say, the mechanism that would be responsible for a peak at 35 mV. We are interested in the general differences that all mechanisms for a peak exceeding 30 mV have in common that the actual mechanism does not have. Whatever they have in common that the actual mechanism does not have, these will be the difference-makers for "30 mV rather than > 30 mV." This procedure generalizes: in order to find the difference-makers for P versus P^*, we have to compare the mechanism M underlying P to the maximally similar mechanism M^* underlying P^*. The difference-makers that explain why P is the case rather than P^* are the differences between M and M* that are shared by all possible mechanisms underlying P*. The explanans for "Why P rather than P*?" will have the form "Because F_1–F_n rather than F^*_1 –F^*_n ", where F_1–F_n and F^*_1 –F^*_n are the difference-makers for P versus P^*.

One consequence of this approach is that the adequacy of the explanans "Because F_1–F_n rather than F^*_1 –F^*_n " is no longer evaluated purely in terms of its accurately describing reality (as it was in the ontic account). Rather, one additional requirement is that the description of the features F_1–F_n has to be specific enough to capture those aspects of the actual mechanism that distinguishes it from all the mechanisms responsible for the contrast phenomenon, and the description of the features F^*_1 –F^*_n must be general enough to capture the similarities between all the mechanisms responsible for the contrast phenomenon.

For example, assume one wants to explain why, between two particular neuronal populations, an action potential propagates rather than not. To answer this question, one has to compare the mechanism that actually underlies the propagation of the action potential between the neuronal populations of interest to the most similar mechanism that would have to occur if the action potential were not propagated. One has to find a description that captures the differences between the actual populations where the action potential is propagated and the most similar mechanism where the action potential is not propagated (which comes down to what is similar between all the possible mechanisms, where the action potential is not propagated). The mechanisms will differ in various respects, such as number of neurons, connections between neurons, size of cell bodies, length of axons, etc. These features will not be mentioned in the answer to the contrastive question. Rather, the answer will contrast more abstract features. Based on such a comparison of mechanisms, the hypothesis has been formulated that the relevant feature that distinguishes between "action potential is propagated" and "action potential is not propagated" is that the neurons in the populations are synchronized, i.e., that the frequency of excitability of the

receiving neurons match with the frequency of activation of the sending neurons (Fries 2015a; 2015b). The answer to the question, "Why does the action potential propagate between these neurons rather than not?" is answered mechanistically: "Because the neurons are synchronized rather than not." This mechanistic explanation is abstract insofar as it does not concern specific structural features of the individual components and their connections.

Thus, Kohár and Krickel's account of mechanistic completeness can account for abstraction. That is, mechanistic explanation no longer has to be given in terms of a biological/chemical/physical entity and activity vocabulary. The difference-makers to be listed in step 4 may be abstract ("abstract" in the sense of "ignoring the specific details of that particular mechanism"). The best way to describe them may not be in terms of the vocabulary of (say) molecular biology but in terms of computational or structural vocabulary.

5.2 How (some) topological explanations are mechanistic explanations

With help of the AD example presented in Section 2.3, we will now show how Kohár and Krickel's account of mechanistic completeness makes sense of the idea that some topological explanations are complete mechanistic explanations.

The central point of the account of mechanistic completeness is that constitutive relevance is a necessary but insufficient condition for showing that some component is explanatorily relevant. Not all constitutively relevant features of a mechanism are relevant to the contrastive explanatory request. This is why Kohár and Krickel's account emphasizes that the explanandum i) needs to take the form of a contrastive question "Why P rather than P^*", and ii) that the explanans is a description of the differences between the mechanism that causes or constitutes the actual phenomenon P and the mechanism that causes or constitutes the contrast phenomenon that would be/have P^*. Thus, the explanans only picks out specific aspects of the actual mechanism. In this account, a mechanistic explanation mentions topological features and is couched in topological terminology (such as "small-worldness" or "centrality") if and only if, for an explanatory request "Why P rather than P^*?", topological terminology provides the best way of capturing the differences between the actual mechanism that brings about P and the mechanisms that would bring about P^* that are at the same time similarities between the mechanisms that would bring about P^*.

That (good) topological explanations can indeed provide the best way of capturing the differences and similarities between the relevant mechanisms will now be illustrated by means of our core example. In line with Kohár and Krickel's account, the explanandum in Yu et al.'s study is, "Why do some AD patients show cognitive deficits and others do not?" The explanans is phrased in terms of topological properties of the brain networks of AD patients in comparison to the brain networks of healthy subjects (which are based on multiplexes modeling the synchronized activity of brain regions within different frequency bands). By first comparing the multiplexes of AD patients with healthy controls in order to assess 'multilayer network disruption' and then correlating the degree of multilayer network disruption with cognitive status within the patients, it could be shown that it is especially those regions normally characterized by high multilayer centrality that are dysfunctional in AD patients. Moreover, the dysfunction levels of multilayer centrality correlate with cognitive deficits in AD patients.

The answer to the explanandum question provided by Yu et al. is "Because, in AD patients with cognitive deficits, interlayer hubs are damaged to a higher degree than in AD patients without cognitive deficits rather than not.' The explanation is topological because "multilayer centrality" is a topological term picking out topological properties of a brain network (see Section 2). Multilayer centrality is multiple realizable by different mechanisms. However, what all the specific mechanisms found in AD patients with different degrees of cognitive deficits have in common is that the regions that correspond to multilayer hub regions in healthy subjects are less connected to other regions compared to the regions found in AD patients without cognitive deficits.

Furthermore, what all the mechanisms found in healthy subjects and AD patients without cognitive deficits seem to have in common is that these specific regions are highly interconnected. Thus, "multilayer centrality" does pick out the similarities between the mechanisms found in AD patients with cognitive deficits which at the same time are i) the differences between the mechanisms found in AD patients and the mechanisms in AD patients without cognitive deficits, and ii) the similarities between the mechanisms in AD patients without cognitive deficits and healthy subjects. In short: the topological explanation of cognitive deficits in AD patients in terms of damage to multilayer hub regions is a mechanistic explanation because "damage to multilayer hub regions" *maps* onto the explanatorily relevant local organizational properties of the brains of all AD patients with cognitive deficits by picking out the explanatorily relevant similarities between the different brains.

Note that Yu and colleagues do not provide a standard topological explanation as the dependency relation between the explanandum and the explanans in this example, which is empirical rather than mathematical: network disruption (decrease of multilayer centrality) is hypothesized to *cause* cognitive decline in AD. Even though the decrease of multilayer centrality was calculated as a mathematical property of the network topology, more empirical information is required to confirm that network disruptions indeed cause the progression of cognitive deficits in patients with AD. Still, the example shows—in line with Ross' argument—that topological explanations can be causal explanations.

6. Conclusion

The aim of this article was to clarify the conditions under which topological explanations are complete mechanistic explanations by critically engaging with Zednik's account of topological explanations as mechanism sketches. As we have shown, the most crucial problem for Zednik's account is that it ignores the difference between global organizational properties and local organization properties and that the former are multiply realizable by the latter. As commonly understood and as captured by the revised mutual manipulability account (MM-O*), mechanistic explanantia can only refer to local organizational properties of mechanistic components. Based on a recent account of mechanistic completeness proposed by Kohár and Krickel (2021), we showed that topological properties can still be parts of mechanistic explanations. We illustrated this by means of the multiplex explanation of cognitive deficits in AD patients.

If successful, the account proposed in this article may be used to solve a general issue surrounding the mechanistic account. The problems afflicting Zednik's account are general problems for the mechanistic account: one common assumption of the new mechanists is that prima facie non-mechanistic explanations must satisfy the *model-to-mechanism-mapping* (3M) requirement (Kaplan 2011), i.e., the variables mentioned in an explanation must map onto working parts of a mechanism, and the dependencies between them must map onto causal connections. This requirement creates a tension with the abstractness of many apparently non-mechanistic explanations whose variables and connections are multiply realizable by different entities and activities and mechanistic-organizational features. The account proposed here can be regarded as the starting point for a general account of the mapping relation between mechanisms and apparently non-mechanistic explanatory descriptions.

Acknowledgements

We thank three anonymous reviewers for their constructive comments. Linda received funding from The Netherlands organization for Scientific Research/NOW (Vidi 198.015).

References

Alstott, Jeffrey, Michael Breakspear, Patric Hagmann, Leila Cammoun, and Olaf Sporns. 2009. "Modeling the Impact of Lesions in the Human Brain." *PLoS Computational Biology* 5 (6). https://doi.org/10.1371/journal.pcbi.1000408.

Baetu, Tudor M. 2012. "Filling in the Mechanistic Details: Two-Variable Experiments as Tests for Constitutive Relevance." *European Journal for Philosophy of Science* 2 (3): 337–53. https://doi.org/10.1007/s13194-011-0045-3.

Baumgartner, Michael, and Lorenzo Casini. 2017. "An Abductive Theory of Constitution." *Philosophy of Science* 84 (2): 214–33. https://doi.org/10.1086/690716.

Baumgartner, Michael, Lorenzo Casini, and Beate Krickel. 2020. "Horizontal Surgicality and Mechanistic Constitution." *Erkenntnis* 85 (2): 417–30. https://doi.org/10.1007/s10670-018-0033-5.

Baumgartner, Michael, and Alexander Gebharter. 2016. "Constitutive Relevance, Mutual Manipulability, and Fat-Handedness." *The British Journal for the Philosophy of Science* 67 (3): 731–56. https://doi.org/10.1093/bjps/axv003.

Baxter, G. J., Dorogovtsev, S. N., Goltsev, A. V., & Mendes, J. F. F. (2012). Avalanche Collapse of Interdependent Networks. *Physical Review Letters*, *109*(24), 248701. https://doi.org/10.1103/PhysRevLett.109.248701.

Boccaletti, S., G. Bianconi, R. Criado, C. I. del Genio, J. Gómez-Gardeñes, M. Romance, I. Sendiña-Nadal, Z. Wang, and M. Zanin. 2014. "The Structure and Dynamics of Multilayer Networks." *Physics Reports* 544 (1): 1–122. https://doi.org/10.1016/j.physrep.2014.07.001.

Craver, Carl F. 2007a. "Constitutive Explanatory Relevance." *Journal of Philosophical Research* 32 (Section II): 1–20. https://doi.org/10.5840/jpr_2007_4.

———. 2007b. *Explaining the Brain: Mechanisms and the Mosaic Unity of Neuroscience*. New York: Oxford University Press.

———. 2014. "The Ontic Account of Scientific Explanation." In *Explanation in the Special Sciences: The Case of Biology and History*, edited by Marie I Kaiser, Oliver R Scholz,

Daniel Plenge, and Andreas Hüttemann, 27–52. Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-007-7563-3_2.

———. 2016. "The Explanatory Power of Network Models." *Philosophy of Science* 83 (5): 698–709. https://doi.org/10.1086/687856.

Craver, Carl F., Stuart Glennan, and Mark Povich. 2021. "Constitutive Relevance & Mutual Manipulability Revisited." *Synthese*, no. 0123456789. https://doi.org/10.1007/s11229-021-03183-8.

Craver, Carl F., and David M. Kaplan. 2020. "Are More Details Better? On the Norms of Completeness for Mechanistic Explanations." *The British Journal for the Philosophy of Science* 71 (1): 287–319. https://doi.org/10.1093/bjps/axy015.

Domenico, Manlio De. 2017. "Multilayer Modeling and Analysis of Human Brain Networks." *GigaScience* 6 (5): 1–8. https://doi.org/10.1093/gigascience/gix004.

Domenico, Manlio De, Albert Solé-Ribalta, Emanuele Cozzo, Mikko Kivelä, Yamir Moreno, Mason A. Porter, Sergio Gómez, and Alex Arenas. 2014. "Mathematical Formulation of Multilayer Networks." *Physical Review X* 3 (4): 1–15. https://doi.org/10.1103/PhysRevX.3.041022.

Fries, Pascal. 2015a. "Communication Through Coherence (CTC 2.0)." *Neuron* 88 (1): 220–35. https://doi.org/10.1016/j.neuron.2015.09.034.Rhythms.

———. 2015b. "Rhythms for Cognition: Communication through Coherence." *Neuron* 88 (1): 220–35. https://doi.org/10.1016/j.neuron.2015.09.034.

Glennan, Stuart. 2017. *The New Mechanical Philosophy*. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780198779711.001.0001.

Harinen, Totte. 2018. "Mutual Manipulability and Causal Inbetweenness." *Synthese* 195 (1): 35–54. https://doi.org/10.1007/s11229-014-0564-5.

Honey, Christopher J., and Olaf Sporns. 2008. "Dynamical Consequences of Lesions in Cortical Networks." *Human Brain Mapping* 29 (7): 802–9. https://doi.org/10.1002/hbm.20579.

Huneman, Philippe. 2010. "Topological Explanations and Robustness in Biological Sciences." *Synthese* 177 (2): 213–45. https://doi.org/10.1007/s11229-010-9842-z.

Illari, Phyllis McKay, and Jon Williamson. 2011. "Mechanisms Are Real and Local." In *Causality in the Sciences*, 818–44. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199574131.003.0038.

———. 2012. “What Is a Mechanism? Thinking about Mechanisms across the Sciences.” *European Journal for Philosophy of Science* 2 (1): 119–35. https://doi.org/10.1007/s13194-011-0038-2.

Jones, David T., David S. Knopman, Jeffrey L. Gunter, Jonathan Graff-Radford, Prashanthi Vemuri, Bradley F. Boeve, Ronald C. Petersen, Michael W. Weiner, and Clifford R. Jack. 2016. “Cascading Network Failure across the Alzheimer’s Disease Spectrum.” *Brain* 139 (2): 547–62. https://doi.org/10.1093/brain/awv338.

Kaplan, David Michael. 2011. “Explanation and Description in Computational Neuroscience.” *Synthese* 183 (3): 339–73. https://doi.org/10.1007/s11229-011-9970-0.

Kästner, Lena. 2017. *Philosophy of Cognitive Neuroscience: Causal Explanations, Mechanisms and Empirical Manipulations*. Berlin/Boston: De Gruyter.

Kivelä, Mikko, Alex Arenas, Marc Barthelemy, James P. Gleeson, Yamir Moreno, and Mason A. Porter. 2014. “Multilayer Networks.” *Journal of Complex Networks* 2 (3): 203–71. https://doi.org/10.1093/comnet/cnu016.

Kohár, M. (2023). *Neural Machines: A Defense of Non-Representationalism in Cognitive Neuroscience* (Vol. 22). Springer International Publishing. https://doi.org/10.1007/978-3-031-26746-8

Kohár, Matej, and Beate Krickel. 2021. “Compare and Contrast: How to Assess the Completeness of Mechanistic Explanation.” In *Neural Mechanisms - New Challenges in the Philosophy of Neuroscience*, edited by Fabrizio Calzavarini and Marco Viola, 395–424. Springer. https://doi.org/10.1007/978-3-030-54092-0_17.

Kostić, Daniel. 2018. “The Topological Realization.” *Synthese* 195 (1): 79–98. https://doi.org/10.1007/s11229-016-1248-0.

———. 2020. “General Theory of Topological Explanations and Explanatory Asymmetry.” *Philosophical Transactions of the Royal Society B: Biological Sciences* 375 (1796). https://doi.org/10.1098/rstb.2019.0321.

Krickel, Beate. 2018a. *The Mechanical World*. Vol. 13. Studies in Brain and Mind. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-03629-4.

———. 2018b. “Saving the Mutual Manipulability Account of Constitutive Relevance.” *Studies in History and Philosophy of Science Part A* 68 (April): 58–67. https://doi.org/10.1016/j.shpsa.2018.01.003.

Mucha, Peter J., Thomas Richardson, Kevin Macon, Mason A. Porter, and Jukka-Pekka Onnela. 2010. "Community Structure in Time-Dependent, Multiscale, and Multiplex Networks." *Science* 328 (5980): 876–78. https://doi.org/10.1126/science.1184819.

Prychitko, Emily. 2021. "The Causal Situationist Account of Constitutive Relevance." *Synthese* 198 (2): 1829–43. https://doi.org/10.1007/s11229-019-02170-4.

Rathkopf, Charles. 2018. "Network Representation and Complex Systems." *Synthese* 195 (1): 55–78. https://doi.org/10.1007/s11229-015-0726-0.

Romero, Felipe. 2015. "Why There Isn't Inter-Level Causation in Mechanisms." *Synthese* 192 (11): 3731–55. https://doi.org/10.1007/s11229-015-0718-0.

Ross, Lauren N. 2021. "Distinguishing Topological and Causal Explanation." *Synthese* 198 (10): 9803–20. https://doi.org/10.1007/s11229-020-02685-1.

Silberstein, Michael, and Anthony Chemero. 2013. "Constraints on Localization and Decomposition as Explanatory Strategies in the Biological Sciences." *Philosophy of Science* 80 (5): 958–70. https://doi.org/10.1086/674533.

Stam, Cornelis J. 2014. "Modern Network Science of Neurological Disorders." *Nature Reviews Neuroscience* 15 (10): 683–95. https://doi.org/10.1038/nrn3801.

Tewarie, Prejaas, Arjan Hillebrand, Bob W. van Dijk, Cornelis J. Stam, George C. O'Neill, Piet Van Mieghem, Jil M. Meier, Mark W. Woolrich, Peter G. Morris, and Matthew J. Brookes. 2016. "Integrating Cross-Frequency and within Band Functional Networks in Resting-State MEG: A Multi-Layer Network Approach." *NeuroImage* 142: 324–36. https://doi.org/10.1016/j.neuroimage.2016.07.057.

Vaiana, Michael, and Sarah Feldt Muldoon. 2020. "Multilayer Brain Networks." *Journal of Nonlinear Science* 30 (5): 2147–69. https://doi.org/10.1007/s00332-017-9436-8.

Watts, Duncan J. 2002. "A Simple Model of Global Cascades on Random Networks." *Proceedings of the National Academy of Sciences of the United States of America* 99 (9): 5766–71. https://doi.org/10.1073/pnas.082090499.

Watts, Duncan J, and Steven H Strogatz. 1998. "Collective Dynamics of 'Small-world' networks." *Nature* 393 (June): 440–42. https://www.ncbi.nlm.nih.gov/pubmed/9623998.

Woodward, James. 2011. "Mechanisms Revisited." *Synthese*, no. January (January): 409–27. https://doi.org/10.1007/s11229-011-9870-3.

———. 2013. "Mechanistic Explanation: Its Scope and Limits." *Aristotelian Society*

Supplementary Volume 87 (1): 39–65. https://doi.org/10.1111/j.1467-8349.2013.00219.x.

Yu, Meichen, Marjolein M.A. Engels, Arjan Hillebrand, Elisabeth C.W. Van Straaten, Alida A. Gouw, Charlotte Teunissen, Wiesje M. Van Der Flier, Philip Scheltens, and Cornelis J. Stam. 2017. "Selective Impairment of Hippocampus and Posterior Hub Areas in Alzheimer's Disease: An MEG-Based Multiplex Network Study." *Brain* 140 (5): 1466–85. https://doi.org/10.1093/BRAIN/AWX050.

Zednik, Carlos. 2019. "Models and Mechanisms in Network Neuroscience." *Philosophical Psychology* 32 (1): 23–51. https://doi.org/10.1080/09515089.2018.1512090.