# 8 Topological Explanations

## An Opinionated Appraisal

*Daniel Kostić*

### 1 Introduction

In recent years, there has been a growing interest in the nature of topological explanations in the sciences (Bechtel 2020; Craver 2016; Darrason 2018; Green et al. 2018; Huneman 2017, 2010; Jones 2014; Kostić 2020a, 2020b, 2018; Kostić and Khalifa 2021; Levy and Bechtel 2013; Matthiessen 2017; Rathkopf 2018; Ross 2021). Most of the accounts agree that a topological explanation appeals to topological properties (for instance, graph-theoretical properties of networks representing complex systems) when explaining certain properties or behaviors of real-world systems. This point has been put forward in a series of papers, starting with Huneman (2010), in which he argued that topological explanations provide a genuine alternative to mechanistic explanations as the mainstream paradigm of explanation in life sciences (Bechtel and Richardson 2010; Craver 2007; Machamer, Darden, and Craver 2000). Similar points have been made in Darrason (2018), Kostić (2018), and Rathkopf (2018), who by discussing different aspects of topological explanations in different areas of science, all argue that topological explanations provide a distinct explanatory strategy in which topological properties play a key role. I call this view "autonomism about topological explanations" and its proponents "autonomists." According to this view, there are cases in which topological properties explain independently of any causal or mechanistic considerations.

In subsequent literature, a few philosophical proposals about what makes these network models explanatory have emerged. The goal of this chapter is to assess them by providing an opinionated appraisal. To that effect, I will argue that my view compares favorably with these other conceptions of topological explanations. I will proceed as follows. Section 2 presents my account of topological explanations. In Section 3, I will compare it with other autonomists' accounts of topological explanations. In Section 4, I will defend my account from key criticisms raised by "mechanistic imperialists," who allege that all topological explanations are mechanistic explanations.

DOI: 10.4324/9781003202905-9

## 2 A General Account of Topological Explanations

In my autonomist view, topological explanations describe counterfactual dependencies between topological properties and empirical properties. A network is a set of vertices and edges. Vertices in different areas of science may represent different things. For example, in neuroscience, vertices might represent neurons or brain regions, whereas the edges are synapses or functional connections between the regions. In computer science, they represent computers, routers, or web pages, while edges represent cables between computers and routers or hyperlinks between web pages. In ecological and food networks, vertices might be species and edges are predation relations. The topological properties are obtained by quantifying network properties of systems by using the graph theory.

For example, an explanation of "why is the rate of the spread of infection in a population as it is, rather than slower or faster?" is "because the population is a small-world network." Such an explanation supports counterfactuals such as: had the population been a regular or random network, the spread of infection in it would have been lower.

With these concepts in hand, I can now provide a general account of topological explanations:

*a*'s being *F* topologically explains why *b* is *G* (*a* and *b* are often identical) if and only if:

(T1): *a* is *F* (where *F* is a topological property);
(T2): *b* is *G* (where *G* is an empirical property);
(T3): Had *a* been *F'* (rather than *F*), then *b* would have been *G'* (rather than *G*);
(T4): That *a is F* is an answer to the question why is *a G*.

The first condition, T1, distinguishes what kinds of properties are taken to be explanatory, and in that way determines whether an explanation is topological or of some other kind. For example, instead of topological properties, *F* could describe statistical properties, or geometrical properties, and the resulting explanations would be statistical or geometrical respectively. T2 ensures that *G* is a proper scientific explanandum (i.e., it is a description of an empirical phenomenon), for example, why a disease spreads at a certain rate in a population (Watts and Strogatz 1998). The third condition T3, secures explanatoriness, that is, T3 captures counterfactual dependency relations. In my earlier example, such a counterfactual would have the following form:

> had the topological properties of contagion relations network in a human population been different, the infection wouldn't have spread as quickly.

Finally, the fourth condition provides contextual criteria for using the counterfactual in two explanatory modes, that is, a horizontal or a vertical explanatory mode. The horizontal and vertical modes emerge from different question asking perspectives. Thus, the counterfactual in the horizontal mode takes the following form:

> (T4H) *a* describes a counterfactual dependency between topological properties and empirical properties that are at the same level,

whereas in the vertical mode it takes this form:

> (T4V) *a* describes a counterfactual dependency between topological properties and empirical properties that are at different levels.

The distinction between vertical and horizontal modes could be illustrated by using local and global topological properties. Local topological properties concern only a part of a network, whereas global ones concern topological properties of the whole network. An example of a local topological property is a node degree, which measures the number of edges maintained by a single node. An example of a global property is average path length, which measures how many edges on average must be traversed to reach any node in a network.

The idea about vertical and horizontal explanatory modes could be extended to more levels than the two general ones that I already mentioned. For example, one could conceive of three levels in ecological networks. In such networks, the nodes represent trophic compartments and edges are the carbon "flows" between trophic compartments. Trophic compartments represent organisms with the same ecological roles (e.g., same prey, predators, or the same metabolic rates). The edges represent the carbon flows or carbon transfers between trophic compartments. The flows can be defined as predator-prey relations, respiration, excretion, that is, the flow quantifies the exchange of carbon biomass (grams of organic carbon per square kilometer per year). In this case, the levels are defined relative to edges as opposed to nodes. And so here we could have a flow level between individual trophic compartments, a cycle level in which flows are embedded into multiple loops within a network (such a loop may exist among several trophic compartments), and the global flow at the whole network level (Niquil et al. 2020).

To illustrate this point, consider an example of an ecological network in which scientists are modeling the effect of offshore wind farms (OWF) on marine ecosystems (Nogues et al. 2021). Building a new OWF in a sandy area provides a new habitat for species like mussels, which is called the reef effect. The reef effect can be measured by looking at the topological recycling property at the node and network level. The node-level recycling is defined by the amount of carbon that is produced by one node and that will circulate into closed loops of flows. The global level is defined by the topological recycling of a whole network, that is, the amount of carbon that is produced by the network that will circulate into closed loops of flows.

The explanation-seeking question in the horizontal mode will be:

> How will the reef effect change the recycling property at the individual trophic compartment level?

Here, the explanation takes the horizontal mode, that is, in answers why the introduction of wind turbines changes the biomass of mussels. Due to wind turbines, the mussels will have an additional habitat, which increases the recycling property of their trophic compartment.

On the other hand, the explanation-seeking question in the vertical mode is:

> How will reef effect change the recycling property of the whole network?

As opposed to the previous example, the explanation here takes the vertical mode, that is, in answers why the introduction of wind turbines changes the biomass of the whole network (due to wind turbines, many species in different trophic comportments will have an additional habitat, which increases the recycling property of the network).

Distinguishing vertical and horizontal explanatory modes in topological explanations is important for understanding how different organizational levels of a system are functionally related as well as how exogeneous changes affect each of the levels. For example, vertical and horizontal explanatory modes provide complementary multilevel explanations of the effect of offshore wind farms on the aquatic ecological community.

Now with my account of topological explanations laid out in detail, in the next section I discuss other autonomists' accounts of topological explanations and highlight some of the major points of agreement between them and some of the fundamental problems in one of them.

## 3 Other Autonomists

Others have written on topological explanation too, and their views can be roughly categorized in two groups. First, *autonomists* about topological explanation, such as myself, maintain that topological explanations are a new and distinct kind of explanation (Darrason 2018; Huneman 2010, 2017; Jones 2014; Kostić 2020a, 2020b, 2018; Kostić and Khalifa 2021, 2022; Rathkopf 2018). Opposed to autonomists are *mechanistic imperialists*, who deny that topological explanations are a distinct kind of explanations and try to subsume topological under mechanistic explanations (Bechtel 2020; Craver 2016; DiFrisco and Jaeger 2019; Glennan 2017; Levy and Bechtel 2013; Matthiessen 2017) or deny their explanatoriness altogether (Craver 2016; Craver and Povich 2017; Povich 2018, 2021). This and the next section focus on autonomists, and subsequent ones discuss mechanistic imperialists.

Most autonomists focus on the distinguishing features of topological explanation. However, apart from Huneman (2017), a few develop any account of topological explanations' structure or source of explanatory power. Since I consider myself an autonomist too, I agree with almost all the points that my fellow autonomists make. These points are not covered in my own work, and hence I see the work of all autonomists as largely complementary.

For example, Rathkopf (2018) argues that mechanistic and topological explanations can be distinguished by their unique explananda, that is, topological explanations are normally used in nearly non-decomposable systems, whereas mechanistic explanations are typically used in decomposable systems. Further, as opposed to mechanistic explanations, topological explanations are actually fueled by the complexity in a system, that is, the very properties that make a system complex are the ones that are explanatory relevant.

Darrason (2018) also argues, just like Rathkopf, that topological explanations have unique explananda, for example, in network medicine they are often used to understand robustness and functional redundancy in certain diseases. She claims that this broader and more general perspective that the topological explanation brings is particularly well suited for distinguishing between monogenetic and polygenetic diseases, which on its own should be a major methodological feat. The uniqueness of *explananda* in topological explanations naturally fits well with the idea that mechanistic and topological explanations are complementary rather than competitors, that is, if topological and mechanistic explanations have different explananda, they cannot compete in the first place.

Others also claim that topological explanations are frequently more abstract than mechanistic explanations (Darrason 2018; Huneman 2010, 2017). Huneman (2010) also argues that apart from topological explanations having different *explananda* in terms of different kinds of physical phenomena, they are almost exclusively used to explain certain general properties of a system, but never to capture its dynamics or some more specific properties and behaviors. But this claim is not warranted by the science literature. As I have shown in various papers (Kostić 2020a, 2020b, 2018; Kostić and Khalifa 2021, 2022), topological explanations are very often used to capture the brain's (temporal) dynamics, or often some local and specific properties of various systems.

Since autonomism claims that some topological explanations are distinct from other kinds of explanations, I also see as autonomists the authors who do not necessarily develop positive arguments for this view but whose views are consistent with it. These authors focus on other epistemic achievements of topological explanations. For example, Green and colleagues (Green et al. 2018) and Serban (2020) argue that topological models frequently provide useful heuristics for discovering mechanisms, but they do not argue that this is the sole function of topological models. As such, their view is compatible with both autonomism and its denial. De Boer and colleagues (2021) examine the applicability of topological explanations in understanding psychiatric disorders.

Ross (2021) makes a closely related but slightly different point. Ross argues that even though the mainstream examples of topological explanation are non-causal, there are some borderline cases that involve both causal and non-causal topological properties. As Ross argues, topologies can be causal when edges in a graph represent causal interactions between a system's entities. These edges are sometimes connected in sequences that are called causal pathways, for example, cell signaling pathways, metabolic pathways, or ecological pathways. These causal pathways can form a network that represents a map or a web of connections in some domain, sometimes called a "wiring topology," and it is similar to wiring diagrams of electric circuits. Ross' argument for why these topologies are causal turns on the fact that wiring topology captures causal properties that can be intervened upon, because these graphs are directed and weighted, and the graphs' edges are causal relations. However, unlike mechanistic imperialists who deny topological autonomism, Ross is more ecumenical and allows that there could be both causal and non-causal distinctively topological explanations. I am very much in agreement with this pluralist idea. Khalifa and I (Kostić and Khalifa 2022) have even elaborated conditions under which a network model can provide a mechanistic explanation.

Now, since these accounts outline only general features of topological explanations, I see them as not only compatible, but also complementary to my own account of topological explanations. The only other account that provides an analysis of the structure and explanatory power in topological explanations is Huneman's (2017) account of topological explanations as a subspecies of structural explanations. A discussion of his account requires a bit more space, and I turn to it in the next subsection.

### *3.1 Topological Explanation as a Subspecies of Structural Explanations*

Huneman (2017) has also offered a philosophical analysis of topological explanation. In what follows, I will present his view, and then I shall highlight several advantages of my own account.

> Huneman's account of topological explanation is a limiting case of his broader analysis of "structural" explanations. Unlike my

> treatment of topological explanations, Huneman does not provide necessary and sufficient conditions for structural explanations. His clearest formulation is: . . . to the extent that [mathematical properties] are explanatory, [i] such properties do not concern variables that would be directly involved in the representation of the system (e.g., time, length, any measurable magnitude of the system)—rather, they constrain any possible causes/mechanisms to have/display a specific behavior/property/outcome range, provided that the system satisfies a small set of conditions (e.g., being describable by a function in the functional class C*, etc.). . . . Whatever the causes in the system, because of one of such properties (T) their mathematical description must display a specific property P, which is mapped onto the explanandum P* we are interested in. In other words, [ii] were this mathematical proposition untrue, the system would not exhibit the property P*, since notwithstanding the actual mechanisms going on here, there [iii] would be no necessity for its mathematical description to display P. Hence those properties account for the explanandum since they provide us with the reason why any possible mechanism proper to the system will have to behave in a specific way that precisely includes having P*; to this extent, no particular mechanism could by itself give us this reason.
>
> (Huneman 2018, 686–687)

Since I am only concerned with topological explanations, I restrict myself to cases in which "*T*" is a topological property. For example, Huneman (2017, 693) offers the following topological explanation as one such example of his structural account of explanation:

- *Explanans:* Small-world networks are stable when random vertices are removed.

  Ecosystems are small-world networks, with species as their vertices.
- *Explanandum:* Ecosystems are stable when random species go extinct.

Here, the major premise in the *explanans* is a mathematical theorem with small-worldliness serving as the topological property *T*, and stability when random vertices are removed serving as Huneman's *P*. Ecosystems' stability when random species go extinct is Huneman's *P**.

Having presented Huneman's position, I will now argue that my view avoids problems that each of the three conditions numbered above—concerning representation, counterfactuals, and necessity, respectively—raise. Further, as the quotation makes clear, these three conditions underwrite Huneman's claim that topological explanations are not mechanistic. This section concludes by showing how the problems with

Huneman's account of structural explanation undercut his argument that topological explanations are non-mechanistic.

However, before proceeding, a small caveat is in order. Since Huneman does not state the precise scope of his account of topological explanations, my discussion of his view can be interpreted in two ways. If Huneman takes all topological explanations to be structural, then the arguments below are direct challenges to his view. If he does not, then my view provides a more comprehensive treatment of topological explanations, including those that are not structural in his intended sense. I remain neutral on which of these two interpretations is correct.

Turning to my criticisms of Huneman's view, consider condition [i]: that topological explanations "do not concern variables that would be directly involved in the representation of the system." Specifically, Huneman's account of structural explanation requires a "mathematical hierarchy," which he characterizes as follows:

> Calling "representative variables" those variables that stand in the model for the properties of the system—e.g., charge, mass, species abundance, activation/nonactivation (of a gate or a neuron), etc.—one can construe a hierarchy of variables: "metavariables" denoting operations on representative variables, "metametavariables" denoting operations on metavariables, etc. Any operation on variables can in turn be part of a set of operations, and as such be represented by a metavariable when this set is described, and so on.
>
> (Huneman 2018, 688)

Huneman takes structural explanations to require metavariables. However, this requirement is difficult to evaluate. For instance, consider a "topological" variable that assumes the value 0 if the ecosystem is small-world, 1 if it is random, and 2 if it is regular. Since this simply encapsulates information about predator-prey relations, it would appear to be a representative variable on Huneman's account. Moreover, if this topological variable did not represent predator-prey relations, then it is hard to see how the graph theory applies to ecosystems in this example, or, more generally, how the graph theory applies to any empirical system. Indeed, some of Huneman's other remarks exacerbate this worry:

> . . . metavariables do not refer to properties of the system (since they are not representative variables). "Reference" here concerns the question whether mathematical properties are directly mapped onto the systems properties—like representative variables—or not.
>
> (Huneman 2018, 688; emphasis in original)

By contrast, my account requires no special class of metavariables and thereby avoids these problems.

Turn now to condition [ii]: that structural explanations work chiefly by supporting mathematical counterpossibles, that is, counterfactuals with antecedents in which the relevant mathematical propositions are false (Huneman 2018, 675). Thus, in the explanation of ecosystems' stability, Huneman requires the following counterfactual to be true:

> Had small-world networks been unstable when random vertices are removed, then the ecosystem would have been unstable if random species were to go extinct.

Since this counterfactual's antecedent requires parts of graph theory to be false, it is a counterpossible. As is well known, such counterpossibles are semantically fraught (Nolan 2013). Indeed, some hold that since no possible worlds exist in which mathematical statements are false, counterpossibles are vacuously true. I consider it a virtue of my view that it does not require counterpossibles. Rather, my account only requires the ecosystem to be capable of having a different topological structure than it does, for example,

> Had the ecosystem not been a small-world network, then it would have been unstable if random species were to go extinct.

This is a garden-variety counterfactual. Assume, as Huneman does, that edges represent predation relations. Then, the counterfactual ranges over possible worlds in which, for example, the ecosystem has different predator-prey relations but the graph theory remains unchanged. Thus, my account requires no mathematical counterpossibles. Indeed, the chief paper from which Huneman draws this example (Montoya, Pimm, and Solé 2006) accords better with my account, for it considers the effects of removing one or more species from the ecosystem, but never considers what would happen if the graph theory were untrue. Thus, *contra* Huneman, the explanation here does not involve counterpossibles. Rather, it involves an empirically testable counterfactual, as my account proposes.

Next, consider condition [iii]: that topological explanations confer an especially strong form of necessity—"in a stronger modal way than the laws of physics" (Huneman 2018, 690)—on their explananda. However, not all topological explanations exhibit extraordinary modal strength. In the ecological example, it is possible for the topological recycling property to be higher or lower, but those possibilities are not results of some mathematical modal constraint (e.g., certain axioms in the graph theory being false); they simply depend on different rates of carbon flows between trophic compartments. Thus, in this case, the explanation works because of the counterfactual dependency between topological recycling and reef effect, and not because of graph-theoretic axioms. It might be thought that my account thereby fails to capture explanations involving high-grade necessity. This is a mistake. Let us follow Huneman in assuming that the explanation of the ecosystem's stability involves this "high-grade necessity." Then we can characterize the explanandum as follows:

> Why is it mathematically impossible that the ecosystem is unstable when new OWFs are built?

Admittedly, because this describes a mathematical impossibility, it is not obvious that this is an empirical statement. Hence, this might imperil satisfaction of T2. However, I assume that because it also describes an ecosystem, treating it as empirical in a broad sense is defensible. Thus construed, T3 thereby suggests that the topological explanation supports the following counterfactual:

> Had the ecosystem not been a network with topological recycling property F, then it would have been mathematically possible that the ecosystem is unstable when new OWFs are built.

Since anything physically possible is also mathematically possible, this counterfactual is true. Hence, my account covers topological explanations that exhibit mathematical necessity.

Thus, my view avoids three problems plaguing Huneman's account. First, my view does not appeal to the problematic notion of metavariables. Second, unlike Huneman's view, mine avoids the notorious semantic puzzles that counterpossibles raise. Third, it covers topological explanations that confer mathematical necessity upon their explananda, as well as those that do not.

Moreover, these three problems undercut Huneman's argument that topological explanations are not a species of causal-mechanical explanation. If no strong distinction between representative variables and metavariables exists, then topological explanations might well be highly abstract causal-mechanical explanations. If topological explanations do not support counterpossibles, then causal-mechanical structure might underwrite the garden-variety counterfactuals they support. Finally, if topological explanations are not restricted to explaining necessities, then causal-mechanical differences—even if represented topologically—might account for when a contingent state of affairs does and does not obtain. Hence, closer scrutiny of Huneman's position undercuts his reasons to think that topological explanations are autonomous.

In the following section, I will provide independent reasons for thinking that topological explanations are autonomous from mechanisms.

## 4 Critiques of Autonomism

Autonomists' views have been challenged in the literature by the proponents of mechanistic explanations in two important ways. First, Craver (2016), Craver and Povich (Craver and Povich 2017), and Povich (Povich 2018, 2021) have argued that topological explanations do not model the right kind of stuff, and that without some ontic backing they are not really explanatory. The second mechanist objection is that networks are insufficiently distinct from mechanisms, especially in terms of organization, and thus if anything they are not a distinct kind of topological model, but merely a very abstract kind of mechanistic model (Bechtel 2020; DiFrisco and Jaeger 2019; Glennan 2017; Levy and Bechtel 2013; Matthiessen 2017). Given this, if they do provide any kind of explanation, it is a mechanistic one.

Khalifa and I (2022) have addressed these objections in much more detail, and hence I will only outline them here. Let me start with the mechanists' interpretation of network topology as mechanistic organization. To equivocate the topological properties of a network with the "organization" in the sense that the organization itself explains (or that the organization of a "mechanism" is responsible for the occurrence of the explanandum) means that topological properties can explain when all other aspects of the mechanism remain fixed. Hence, if topology alone is explanatory, then it is obviously a distinctively topological explanation. This either concedes precisely what is at stake or it trivializes the mechanistic conception of "organization." And second, to say that topology as the organization of a mechanism does not explain is to beg the question against the autonomist.

To develop these arguments in more detail, Khalifa and I proposed an analysis that we call a Mechanistic Interpretation of Topological Explanations or a MITE. According to MITE, a network model provides a mechanistic explanation, if and only if network nodes represent a mechanism's entities and activities; edges represent the interactions between entities and activities; the network model specifies how these entities and activities are organized to be responsible for a target phenomenon; and finally, the target phenomenon is at a higher level than the mechanism's entities and activities. If only one of these conditions is violated, the MITE fails, and the resulting explanation is not mechanistic. This, in itself, does not imply that the resulting explanation is distinctively topological either. We provided examples in which at least one of the MITE's requirements have been violated, but the explanation satisfied T1–T4, which means that the resulting explanation is topological and non-mechanistic (Kostić and Khalifa 2022).

In addition, mechanists who subscribe to ontic conceptions of explanation have provided an additional challenge to autonomism (Craver and Povich 2017). They claim that without ontic backing, topological explanations fail to account for directionality, that is, they incorrectly deem instances of logical contraposition as correct explanations. According to the directionality problem, topological models without ontic backing are bidirectional, yet no bidirectional model is explanatory. So, purveyors of this objection conclude that no topological models without ontic backing are explanatory. In response, Khalifa and I (2021) provide a more precise definition of the directionality problem, distinguish it from the asymmetry problem, and formulate what we call the "ontic irrelevance lesson." The ontic irrelevance lesson uses the ideas from my previous work (Kostić 2020a) to show that topological explanations can preserve directionality without any kind of ontic backing that Craver (2016), Craver and Povich (2017), and Povich (2018, 2021) require. Non-ontic bases of directionality of topological explanations that we proposed are the property (which is tied to T1 and T2), counterfactual (T3), and perspectival ones (T4) (Kostić and Khalifa 2021).

Khalifa and I argue (2021) that a topological explanation will be property directional when its contraposition or "reversal" does not involve topological properties in their "explanantia." For example, Helling and colleagues (Helling, Petkov, and Kalitzin 2019) offer a topological explanation in which mean functional connectivity[1] explains the dynamics of the onset of epileptic seizure (also called *ictogenicity*). If the reversal of their original case is an explanation, then ictogenicity must be a topological property. However, it is not. Hence, Helling et al.'s actual explanation satisfies the property requirement, but its reversal does not.

In counterfactual directionality, the original case satisfies a counterfactual, but its reversal does not. Using the study by Adachi and colleagues (Adachi et al. 2012), which shows that the functional connectivity between anatomically unconnected areas in the macaque cortex counterfactually depends on the anatomical connectivity network's overall frequency of three-node network motifs, Khalifa and I showed that a reversal of this counterfactual violates the very dependency that the original case postulates. Thus, if the relevant counterfactual is true, its reversal is false, and this difference alone suffices to capture the directionality of the explanation. Thus, it exhibits *counterfactual directionality*.

Finally, we argued that topological explanations must be answers to the relevant explanation-seeking questions, which is embodied in the T4-Perspectival condition. Thus, an explanation is *perspectivally directional* whenever its original satisfies T4 condition, but its reversal does not. Explanation-seeking questions allow that in the same context *at least one* answer will appeal to a topological property, but its reversal will likely *prohibit* any topological answers.

This is another aspect in which my account of topological explanations outperforms Huneman's. Namely, Huneman's example states:

1 *Explanans:* Small-world networks are stable when random vertices are removed.

2 Ecosystems are small-world networks, with species as their vertices.
3 *Explanandum:* Ecosystems are stable when random species go extinct.

However, this is bidirectional, because it can be reversed in the following way, and Huneman's account would incorrectly classify it as a correct explanation:

1 *Explanans:* Small-world networks are stable when random vertices are removed.
2 Some ecosystems are *unstable* when random species go extinct.
3 *Explanandum:* Some ecosystems are either *not* small-world networks or do *not* have species as their vertices.

So, this poses an additional problem with Huneman's account: its vulnerability to the directionality problem.

Two points remain to be made in defense of topological autonomism. Autonomism argues that some topological explanations are non-mechanistic. Moreover, I have claimed that T1–T4 provide sufficient conditions for genuinely autonomous topological explanations. However, Craver (2016, 704–706) argues that functional connectivity models are examples of topological models that are not explanations, chiefly because they do not represent mechanisms. Functional connectivity measures temporal dependency between time series of anatomically separated brain regions (van den Heuvel and Hulshoff Pol 2010, 519). In functional connectivity models, the nodes represent blood-oxygen-level-dependent (BOLD) signals (in fMRI) or EEG channels (in EEG recordings) and the edges represent synchronization correlations between BOLD signals or EEG channels. The idea is that if two BOLD signals (or EEG channels) are synchronized, the populations of neurons that the BOLD/EEG signals represent are functionally connected. Craver observes that functional connectivity models' nodes "need not . . . stand for working parts," that is, for the entities that constitute a mechanism. Rather, many functional connectivity models' nodes are conventionally determined spatiotemporal regions that are adopted mostly because they are "conveniently measurable units of brain tissue rather than known functional parts."

The spearpoint of his argument is that models representing correlations are merely evidential, but not explanatory (Craver 2016, 705). This argument turns on the idea that if nodes are not working parts of a mechanism and the edges are merely correlations, then functional connectivity networks do not model the right kinds of stuff, and because of that they cannot be explanatory. Just like the barometer reading is correlational evidence of a storm, but it does not explain the storm, the synchronization likelihoods in functional connectivity networks are evidence of the actual anatomical connections and BOLD/EEG signals are evidence of neuronal populations. To model the right kinds of stuff and be explanatory, synchronization likelihoods and BOLD/EEG signals would have to be working parts of a mechanism that is responsible for the phenomenon we want to explain. Since they are not, then functional connectivity models cannot be explanatory, according to Craver. This is the core of Craver's challenge.

Responding to Craver's challenge allows me to sketch what I will call "ambitious autonomism." According to this view, I can grant that in general, most models trafficking only in correlations between conventionally determined spatiotemporal regions are evidential but not explanatory, but I argue that if some functional connectivity models, such as the one discussed in this chapter, satisfy T1–T4 conditions and also violate some or all of the MITEs, then they are exceptions of the Craver's challenge, in that they fail to be mechanistic, but are still explanatory. Ambitious autonomism might draw inspiration from the fact that some functional connectivity networks capture pathological brain dynamics that structural connectivity (i.e., mechanistic structure) cannot. This is because variability in functional connectivity often occurs in the absence of direct anatomical connections (Helling, Petkov, and Kalitzin 2019; Honey et al. 2009; Moon et al. 2017; Suárez et al. 2020). This is a *prima facie* methodological reason to countenance functional connectivity explanations that are autonomous of any mechanistic explanation. Recall that functional connectivity edges are synchronization likelihoods. Thus, many of the functional connectivity models' topological properties are different ways of describing how well synchronized a brain is. For instance, the fact that the likelihood of an epileptic seizure (or ictogenicity) varies in proportion to mean functional connectivity suggests that "oversynchronization" of the brain explains seizures (Kalitzin et al. 2019, 7). Thus, if the brain dynamics counterfactually depends on variability in functional connectivity, whereas the direct anatomical connections remain fixed, then such a model satisfies T1–T4. This shows that functional connectivity models can be explanatory, and they are not always merely evidential.

In addition, recent philosophical work at the intersection of modeling and explanation further dispels this intuition in the case of explanatory functional connectivity models. Ambitious autonomists may interpret functional connectivity models as showing that the specific mechanistic entities and activities simply *don't matter* when it comes to explaining certain phenomena. In the case of ictogenicity, oversynchronization (i.e., high mean functional connectivity) suffices on its own. As such, functional connectivity models' representation of brain regions as arbitrary spatiotemporal regions is a kind of explanatory abstraction (Jansson and Saatsi 2017) or idealization in which mechanistic entities are caricatured as arbitrary spatiotemporal regions to highlight their explanatory irrelevance (Batterman and Rice 2014).

## 5 A Conclusion

Topological explanation is a recent scientific development, which by now has received significant philosophical attention. It already has some devoted proponents, autonomists, but it also has staunch critics from the ranks of new mechanistic philosophers. Among the autonomists, there is a broad agreement about some general features of topological explanation, which make them a unique and a novel kind of explanation. Such features are that they are often invariant relative to causal or microphysical details of real-world systems, they are fueled rather than hindered by the complexity of a system, they are particularly well suited for explaining nearly non-decomposable systems, they are often used to understand robustness and functional redundancy in certain systems, they are also more abstract than their mechanistic counterparts, and they are often non-causal. When it comes to their structure, there is also a broad agreement about certain features, for example, in contrast to traditional accounts of realization relation, the topological realization base is normally at a higher or the same level as the realized properties (Huneman 2018; Kostić 2018). However, when it comes to the precise analyses of topological explanation, there are two developed accounts: mine and Huneman's. As I have argued, I see my account as enjoying several advantages, such as avoiding problems with metavariables, counterpossibles, the problems with mathematical necessity, directionality problems, as well as being in closer dialogue with the scientific literature.

The critics of topological explanation have pointed out some really important issues that directly challenge the autonomy of topological explanations, for example, the ontic backing, the directionality, or the representational power of network models. However, the autonomists have ample resources to meet these challenges head on and disarm them in many different ways (Kostić 2020a, 2020b, 2018; Kostić and Khalifa 2021, 2022). Finally, autonomists are more ecumenical. For example, we argue that topological explanations can come in many different flavors, e.g., both causal (Ross 2021) and non-causal or distinctively topological and mechanistic (Kostić and Khalifa 2022); or that topological models can have many important non-explanatory uses, such as providing heuristics for discovering mechanisms or parts of mechanisms (de Boer et al. 2021; Green et al. 2018; Serban 2020).

This diversity of perspectives and views illustrates the richness and a broad potential that topological explanations bring to the table. Current accounts and debates about them are only the beginning of a fruitful development in the philosophy of scientific explanation.

## Acknowledgments

I would like to thank Kareem Khalifa for his invaluable comments and insights that helped bring my work on topological explanations in general, but also specifically in this chapter, in much greater focus. I would also like to thank Kareem for the wonderful friendship.

## Funding

I would like to acknowledge the funding by the Radboud Excellence Initiative.

## Note

1 Which measures the average strength of the correlations that exist between any two nodes in a functional connectivity network.

## References

Adachi, Yusuke, Takahiro Osada, Olaf Sporns, Takamitsu Watanabe, Teppei Matsui, Kentaro Miyamoto, and Yasushi Miyashita. 2012. “Functional Connectivity between Anatomically Unconnected Areas Is Shaped by Collective Network-Level Effects in the Macaque Cortex.” *Cerebral Cortex* 22 (7): 1586–1592. https://doi.org/10.1093/cercor/bhr234.

Batterman, Robert W., and Collin C. Rice. 2014. “Minimal Model Explanations.” *Philosophy of Science* 81(3): 349–376. https://doi.org/10.1086/676677

Bechtel, William. 2020. “Hierarchy and Levels: Analysing Networks to Study Mechanisms in Molecular Biology.” *Philosophical Transactions of the Royal Society B: Biological Sciences* 375 (1796): 20190320. https://doi.org/10.1098/rstb.2019.0320.

Bechtel, William, and Robert C. Richardson. 2010. *Discovering Complexity: Decomposition and Localization as Strategies in Scientific Research*. MIT Press ed. Cambridge, MA: MIT Press.

Craver, Carl F. 2007. *Explaining the Brain: Mechanisms and the Mosaic Unity of Neuroscience*. Oxford: Clarendon Press.

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

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.

Darrason, Marie. 2018. “Mechanistic and Topological Explanations in Medicine: The Case of Medical Genetics and Network Medicine.” *Synthese* 195 (1): 147–173. https://doi.org/10.1007/s11229-015-0983-y.

de Boer, Nina S., Leon C. de Bruin, Jeroen J. G. Geurts, and Gerrit Glas. 2021. “The Network Theory of Psychiatric Disorders: A Critical Assessment of the Inclusion of Environmental Factors.” *Frontiers in Psychology* 12 (February): 623970. https://doi.org/10.3389/fpsyg.2021.623970.

DiFrisco, James, and Johannes Jaeger. 2019. "Beyond Networks: Mechanism and Process in Evo-Devo." *Biology & Philosophy* 34 (6): 54. https://doi.org/10.1007/s10539-019-9716-9.

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): 1751–1777. https://doi.org/10.1007/s11229-016-1307-6.

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." In *Proceedings of the 2nd International Conference on Applications of Intelligent Systems*, 1–6. Las Palmas de Gran Canaria Spain: ACM. https://doi.org/10.1145/3309772.3309775.

Honey, Christopher J., Jean-Philippe Thivierge, and Olaf Sporns. 2010. "Can Structure Predict Function in the Human Brain?" *Neuroimage* 52 (3): 766–776. https://doi.org/10.1016/j.neuroimage.2010.01.071

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

———. 2017. "Outlines of a Theory of Structural Explanations." *Philosophical Studies*, no. 1984: 1–38. https://doi.org/10.1007/s11098-017-0887-4.

———. 2018. "Realizability and the Varieties of Explanation." *Studies in History and Philosophy of Science Part A* 68 (April): 37–50. https://doi.org/10.1016/j.shpsa.2018.01.004.

Jansson, Lina, and Saatsi, Juha. 2017. "Explanatory Abstractions". *The British Journal for the Philosophy of Science,* 70 (3): 817–844. https://doi.org/10.1093/bjps/axx016

Jones, Nicholaos. 2014. "Bowtie Structures, Pathway Diagrams, and Topological Explanation." *Erkenntnis* 79 (5): 1135–1155. https://doi.org/10.1007/s10670-014-9598-9.

Kalitzin, Stiliyan, George Petkov, Piotr Suffczynski, Vasily Grigorovsky, Berj L. Bardakjian, Fernando Lopes da Silva, and Peter L. Carlen. 2019. "Epilepsy as a Manifestation of a Multistate Network of Oscillatory Systems." *Neurobiology of Disease* 130: 104488. https://doi.org/10.1016/j.nbd.2019.104488

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

———. 2020a. "General Theory of Topological Explanations and Explanatory Asymmetry." *Philosophical Transactions of the Royal Society B: Biological Sciences* 375 (20190314): 1–8. http://dx.doi.org/10.1098/rstb.2019.0321.

———. 2020b. "Minimal Structure Explanations, Scientific Understanding and Explanatory Depth." *Perspectives on Science 2019* 27 (1): 48–67. https://doi.org/10.1162/posc.

Kostić, Daniel, and Kareem Khalifa. 2021. "The Directionality of Topological Explanations." *Synthese* 199: 14143–14165. https://doi.org/10.1007/s11229-021-03414-y.

Kostić, Daniel and Kareem Khalifa. 2022. "Decoupling Topological Explanations from Mechanisms."

*Philosophy of Science,* 1-39. doi:10.1017/psa.2022.29

Levy, Arnon, and William Bechtel. 2013. "Abstraction and the Organization of Mechanisms." *Philosophy of Science* 80 (2): 241–261. https://doi.org/10.1086/670300.

Machamer, Peter, Lindley Darden, and Carl F. Craver. 2000. "Thinking about Mechanisms." *Philosophy of Science* 67 (1): 1–25. https://doi.org/10.1086/392759.

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

Montoya, José M., Stuart L. Pimm, and Ricard V. Solé. 2006. "Ecological Networks and Their Fragility." *Nature* 442 (7100): 259–264. https://doi.org/10.1038/nature04927

Moon, Joon-Young, Junhyeok Kim, Tae-Wook Ko, Minkyung Kim, Yasser Iturria-Medina, Jee-Hyun Choi, Joseph Lee, George A. Mashour, and UnCheol Lee. 2017. "Structure Shapes Dynamics and Directionality in Diverse Brain Networks: Mathematical Principles and Empirical Confirmation in Three Species." *Scientific Reports* 7 (1): 1–11. https://doi.org/10.1038/srep46606

Niquil, Nathalie, Matilda Haraldsson, Télesphore Sime-Ngando, Philippe Huneman, and Stuart R. Borrett. 2020. "Shifting Levels of Ecological Network's Analysis Reveals Different System Properties." *Philosophical Transactions of the Royal Society B: Biological Sciences* 375 (1796): 20190326. https://doi.org/10.1098/rstb.2019.0326.

Nogues, Quentin, Aurore Raoux, Emma Araignous, Aurélie Chaalali, Tarek Hattab, Boris Leroy, Frida Ben Rais Lasram, et al. 2021. "Cumulative Effects of Marine Renewable Energy and Climate Change on Ecosystem Properties: Sensitivity of Ecological Network Analysis." *Ecological Indicators* 121 (February): 107128. https://doi.org/10.1016/j.ecolind.2020.107128.

Nolan, Daniel P. 2013. "Impossible Worlds." *Philosophy Compass* 8 (4): 360–372. https://doi.org/10.1111/phc3.12027

Povich, Mark. 2018. "Minimal Models and the Generalized Ontic Conception of Scientific Explanation." *The British Journal for the Philosophy of Science* 69 (1): 117–137. https://doi.org/10.1093/bjps/axw019.

———. 2021. "The Narrow Ontic Counterfactual Account of Distinctively Mathematical Explanation." *The British Journal for the Philosophy of Science* 72 (2): 511–543. https://doi.org/10.1093/bjps/axz008.

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

Ross, Lauren N. 2021. "Distinguishing Topological and Causal Explanation." *Synthese* 198 (10): 9803–9820. https://doi.org/10.1007/s11229-020-02685-1.

Serban, Maria. 2020. "Exploring Modularity in Biological Networks." *Philosophical Transactions of the Royal Society B: Biological Sciences* 375 (1796): 20190316. https://doi.org/10.1098/rstb.2019.0316.

Suárez, Laura E., Ross D. Markello, Richard F. Betzel, and Bratislav Misic. 2020. "Linking Structure and Function in Macroscale Brain Networks." *Trends in Cognitive Sciences* 24 (4): 302–315. https://doi.org/10.1016/j.tics.2020.01.008

van den Heuvel, Martijn P., and Hilleke E. Hulshoff Pol. 2010. "Exploring the Brain Network: A Review on Resting-State FMRI Functional Connectivity." *European Neuropsychopharmacology* 20 (8): 519–534. https://doi.org/10.1016/j.euroneuro.2010.03.008.

Watts, Duncan J., and Steven H. Strogatz. 1998. "Collectivedynamics of 'Small-World' Networks." *Nature* 393 (6684): 440–442. https://doi.org/10.1038/30918.