PHILOSOPHICAL TRANSACTIONS B

royalsocietypublishing.org/journal/rstb

Research

Cite this article: Kostić D. 2020 General theory of topological explanations and explanatory asymmetry. *Phil. Trans. R. Soc. B* **375**: 20190321.
http://dx.doi.org/10.1098/rstb.2019.0321

Accepted: 23 September 2019

One contribution of 11 to a theme issue 'Unifying the essential concepts of biological networks: biological insights and philosophical foundations'.

Subject Areas:
neuroscience, behaviour

Keywords:
philosophical theory of explanation, explanatory asymmetries, explanatory perspectivism, counterfactual dependencies, facticity of explanation, explanatory unification

Author for correspondence:
Daniel Kostić
e-mail: daniel.kostic@gmail.com

THE ROYAL SOCIETY PUBLISHING

General theory of topological explanations and explanatory asymmetry

Daniel Kostić

University Bordeaux Montaigne, Department of Philosophy and EA 4574 'Sciences, Philosophie, Humanités' (SPH) at University of Bordeaux, Allée Geoffroy Saint-Hilaire, Bâtiment B2, 33615 Pessac cedex, France

DK, 0000-0001-5729-1476

In this paper, I present a general theory of topological explanations, and illustrate its fruitfulness by showing how it accounts for explanatory asymmetry. My argument is developed in three steps. In the first step, I show what it is for some topological property A to explain some physical or dynamical property B. Based on that, I derive three key criteria of successful topological explanations: a criterion concerning the facticity of topological explanations, i.e. what makes it true of a particular system; a criterion for describing counterfactual dependencies in two explanatory modes, i.e. the vertical and the horizontal and, finally, a third perspectival one that tells us when to use the vertical and when to use the horizontal mode. In the second step, I show how this general theory of topological explanations accounts for explanatory asymmetry in both the vertical and horizontal explanatory modes. Finally, in the third step, I argue that this theory is universally applicable across biological sciences, which helps in unifying essential concepts of biological networks.

This article is part of the theme issue 'Unifying the essential concepts of biological networks: biological insights and philosophical foundations'.

1. Introduction

Even though network-based explanations (hereafter, topological explanations) have been used in sciences for several decades, we do not yet have a theory of topological explanations that can delimit good from bad topological explanations in a principled way. The most general idea of topological explanation is that it describes how the mathematical properties of connectivity patterns in complex networks determine the dynamics of the systems exhibiting those patterns [1]. A network is a collection of nodes and edges that are connected in certain ways; and a graph is a mathematical description of such a network ([2], p. 683). Topological properties are mathematical properties obtained by quantifying networks by using graph theory and similar approaches. Since many biological systems can be modelled as networks, i.e. they have many interconnected elements that can be considered as nodes and edges, this approach clearly has enormous explanatory potential. This fact is even more significant given the sheer microphysical diversity of biological systems, e.g. the same topological explanatory pattern can be used to explain the robustness of the brain, a computer network, an ecological community, a protein interaction network and so on.

One of the major demands on such a theory of topological explanations is to account for what it is for a certain pattern to be explanatory as a matter of a principle (why the explanation succeeds, or why it does not succeed), as well as to account for explanatory asymmetries, i.e. in a good explanation, if A explains B then B should not explain A; otherwise, the explanation is circular or too permissive.

In this paper, I present a general theory of topological explanations and illustrate its fruitfulness by showing how it accounts for explanatory asymmetries. My argument is developed in the three steps. In the first step, I show what it is for some topological property A to explain some physical or dynamical property B (§2). Based on that, I derive three key criteria of successful topological explanation: a criterion concerning the facticity of topological explanations, i.e. what makes it true of a particular system; a criterion for two explanatory modes, i.e. describing what I dub 'vertical' and 'horizontal' counterfactual dependencies and, finally, a third one that tells us when to use the vertical and when to use the horizontal mode. In the second step, I show how this theory of topological explanations accounts for explanatory asymmetry in several different ways (§§3 and 4). Finally, in the third step, I argue that this theory applies across all biological networks (§5).

2. A general theory of topological explanations

A general theory of topological explanations should tell us what it is for a certain network pattern to be explanatory for any particular system or a phenomenon that we want to explain using that pattern. To that end, I characterize topological explanation in the most general sense, and precisely define its structure so as to capture different explanatory modes, i.e. the vertical and the horizontal ones.

Let us start with the most general definition of kind of things a topological explanation is: a topological explanation supports counterfactuals that describe a counterfactual[1] dependency between a system's topological properties and its network dynamics.

There are two different ways in which topological explanations may describe counterfactual dependency relations, i.e. which I call the 'vertical' and the 'horizontal'. By 'vertical', I mean an explanation in which a global topological property determines certain general properties of the real-world system. On the other hand, by 'horizontal' I mean an explanation in which a local topological property determines certain local dynamical properties of the real-world system. Only whole networks possess global properties; only parts of networks possess local properties. Examples of the global–local distinction include the distinctions between within-scale (horizontal) and between-scales (vertical) and intra-modular (horizontal) and hierarchical–modular (vertical) and so on.

To that effect, I propose a definition of topological explanation that is sensitive to both vertical and horizontal approaches:

A topologically explains *B* if and only if:

1. (*Facticity*): *A* and *B* are approximately true; and
2. Either
   (a) (*Vertical mode*): *A* describes a global topology of the network, *B* describes some general physical property, and had *A* had not obtained, then *B* would not have obtained either; or
   (b) (*Horizontal mode*): *A* describes a set of local topological properties, *B* describes a set of local physical properties, and had the values of *A* been different, then the values of *B* would have been different.
3. (*Explanatory perspectivism*): *A* is an answer to the relevant explanation-seeking question Q about *B*, such that the Q determines whether to use the vertical or horizontal explanatory mode.

The facticity requirement is a general condition according to which the *explanans*[2] and *explanandum*[3] have to be approximately true, and it works the same in both vertical and horizontal explanatory modes.[4] In short, correct topological explanations cannot rely on gross misrepresentations of a system's topological properties or physical properties. This point can also be framed in terms of conditions of application, which are elaborated in great detail in another contribution [9] in this theme issue.

Turning to the second condition, one can illustrate the difference between vertical and horizontal modes of topological explanation by looking at how they apply to closely related *explananda*. The brain is able to process information extremely efficiently by engaging in a variety of very complex behaviours, e.g. it performs various cognitive tasks and can go through many different states. To understand some of its dynamics, i.e. how it is possible for it to go through different states (for example, from a resting state, to a state of solving a logical puzzle, or from a state of solving a logical puzzle to a state of retrieving information from memory) we want to understand how it achieves cognitive control, i.e. how the brain as a dynamical system achieves efficient transition from one of its internal states to another. This can be done by looking into what keeps the energy cost of such transitions low.

When explaining cognitive control in either mode, the brain is represented as a network of brain regions connected through white matter tracts. In the vertical mode, the *explanans* is the global topological property such as small-worldliness and the *explanandum* is the global physical property such as global cognitive control (if global cognitive control is theoretically possible). The most basic way to illustrate small-worldliness as a global topological property is through the Watts & Strogatz [10] model. Among the most established ways to quantify networks are the average path length $L(p)$ and clustering coefficient $C(p)$. The $L(p)$ measures the average number of edges that have to be traversed in order to reach one node from the other. Clustering is understood as a tendency of a small group of nodes to form connected triangles around one central node, which indicates that the connected neighbouring nodes are also neighbours among themselves; hence they form a cluster or a clique. The clustering coefficient is a measure of this tendency, which characterizes a value for all nodes in a network ([11], p. 312; [2], p. 683). Networks that have high clustering coefficients and low path lengths are called small-world networks [10]. Small-worldliness as a global topological property indicates that almost any two nodes in the network will be connected either directly or through a minimal number of indirect connections, which shortens the distance between the nodes within a neighbourhood of nodes as well as between neighbourhoods of nodes, and neighbourhoods of neighbourhoods, which further ensures that the energy requirements for changing any of the trajectories will be minimal, and thus explains why the network is globally or in principle controllable.

If the explanation-seeking question is: why is the brain controllable, then the explanation will have to be given in the vertical mode. The relevant 'vertical counterfactual' then is

> Had the brain not been a small-world network it wouldn't have been controllable.

On the other hand, in the horizontal approach, *A* describes a set of local topological properties, *B* describes a set of local physical properties, and had the values of *A* been different, then the values of *B* would have been different. For example, when applied to a particular case of cognitive controllability, the explanation-seeking question is: *How and why is the brain able to efficiently transition from one state to the other?* The answer proposed in a series of studies [12–14] is that local topological properties determine energy requirements for those movements ([13], p. 1). Specifically, in recent literature in cognitive neuroscience [12–15], the problem of cognitive control is being treated through the notion of network control.

It should be emphasized from the outset that these two notions of control are very different. The notion of cognitive control refers to a brain's dynamics, which can be understood as a neural regional activity that can be elicited by neurofeedback in fMRI imaging, or by non-invasive brain stimulation such as TMS (trans-cranial-magnetic stimulation). In this sense, to control means: how to perturb the system in order to reach a desired state. By contrast, the notion of network control is a purely mathematical notion in network control theory which refers to topological constraints on such perturbations.[5] The dependency that the explanation describes in this case is between the topological properties and cognitive control—the idea that the topological properties exert network control over cognitive control understood as an aspect of the brain's dynamics, e.g. certain connectivity patterns enable more cognitive control, while others enable less cognitive control. As we shall shortly see, this difference will have to do with the specific values of local topological properties, e.g. the higher the network communicability[6] value the lower the energy requirements for changing the particular trajectories.

In one of the programmatic papers on this approach in cognitive neuroscience, Gu and colleagues [12] make this distinction explicit:

> Importantly, this notion of control is based on a very detailed mathematical construct and is therefore necessarily quite distinct from the cognitive neuroscientist's common notion of 'cognitive control' and the distributed sets of brain regions implicated in its performance. To minimize obfuscation, we henceforth refer to these two notions as 'network control' and 'cognitive control', respectively. ([12], p. 8)

The distinction between cognitive control and the network control is of great importance here because it helps to clearly delimit the *explanans* from *explanandum* in the horizontal mode of topological explanation, i.e. we want to explain how the cognitive control is achieved by using the tools of network control theory to describe counterfactual dependencies between topological properties and the network dynamics.

The horizontal counterfactual would then be: 'Had the network's communicability been higher, then the energy required to change a trajectory in a state space would have been lower.' When we plug this example into the horizontal explanatory mode condition, we can ascertain that the communicability is a local topological property, energy requirements are local dynamical properties, and the energy requirements counterfactually depend on the communicability measure.

This pattern of counterfactual dependency is explanatory precisely because it tells us how hypothetical changes in the values of topological properties would affect the system dynamics.

This example also highlights key differences between the vertical and horizontal modes of topological explanation. In the horizontal mode, the facticity requirement is satisfied by describing the system dynamics as a state space in which transitions from one state to the next are described as trajectories in a state space. Understanding why some trajectories are controllable has to do with understanding the counterfactual dependency between local topological measures of the network and the particular trajectories in the state space, i.e. the explanation tells us in what ways exactly the trajectories in the state space will be affected by the changes of relevant topological properties. As opposed to cases of vertical counterfactual dependencies, horizontal counterfactual dependencies hold between variables that are at the same local level in the network. Thus, in the horizontal approach, the relationship between the *explanans* and *explanandum* is more direct: just by describing the topological properties of the trajectories in the state space we are almost immediately able to understand the relevant counterfactual dependency relations, without needing to appeal to any kind of inferential patterns between different levels. This indicates that in the horizontal approach the topological explanation has what Kostić [1] calls minimal structure, where the relationship between the *explanans* and *explanandum* is more direct.

The final criterion tells us when to use the vertical and when to use the horizontal mode. The idea, foreshadowed by Achinstein [16] and van Fraassen [17] in more general discussions of explanation, is what Kostić [18] calls explanatory perspectivism: explanation-seeking questions determine the mode (horizontal or vertical) of topological explanations. To appreciate this point, consider the idea recently proposed by Hilgetag & Goulas [19]. They question the very notion of small-world topology as the crucial aspect for understanding the efficiency of the brain organization in signal processing. Given the multitudes of ways in which the small-worldiness can be realized in very diverse arrangements of topological properties, they argue that, if each of the arrangements of topological properties that can be used to describe the small-world topology describes a different pattern of dependencies between topological structure and the dynamical features in the brain, then the global notion of small-world topology does not seem very informative. The truly explanatory dependencies should be the ones that tell us various hypothetical ways in which the changes in the local values of topological variables would affect the properties and behaviours of dynamical elements such as trajectories in the state space. However, this does not imply that the vertical approach is explanatorily superfluous. Rather, it suggests that explanation-seeking questions dictate whether to use the vertical or horizontal approach. For example, if we want to explain some very general property of the system or of a phenomenon, e.g. if a system is globally controllable at all or why it is stable at its most global level, we will use the vertical approach, because it gives us a very general answer. However, if we want to know why the system is controllable in particular ways, we will use the horizontal approach. As illustrated above, it is possible to use both the vertical and horizontal explanatory modes to explain the same phenomenon, depending on what we want to know about the phenomenon.

The theory of topological explanations developed in this section thus provides three criteria: a criterion of what makes the explanation true of a particular system; a clear distinction between two different modes of topological explanations, the vertical and the horizontal modes; and a criterion for when to use one or the other modes of topological explanations.

3. Background of the asymmetry problem

In this section, I provide some background of the asymmetry problem. This sets the stage for how the general theory of

logical deduction	$C_1, C_2, \ldots, C_k$	statements of antecedent conditions	*explanans*
	$L_1, L_2, \ldots, L_r$	general laws	
	E	description of the empirical phenomenon to be explained	*explanandum*

Figure 1. Hempel and Oppenheim ([20], p. 138).

topological explanations developed above accounts for the asymmetry problem, and how it also helps to unify essential concepts of biological networks in the remainder of the paper.

Beginning in the second part of the twentieth century, one of the central topics in the philosophy of science was the scientific explanation. Several very sophisticated philosophical accounts of scientific explanation have emerged, each of which proposed a set of universal epistemic norms to govern any successful scientific explanation.

The first and most influential account was the deductive-nomological model (DN model hereafter) developed by Hempel & Oppenheim [20] and Hempel [21]. They argued that the explanation has an argument structure, in which the *explanandum* is the conclusion in the logical argument that is derived from a set of premises that constitute the *explanans*. The set of premises in the *explanans* are constituted, respectively, by the statements describing the antecedent conditions and the statements describing the general laws of nature, and the only constraining conditions are that the statements in the *explanans* must be empirically true and that the deductive argument must be valid. Famously, Hempel & Oppenheim [20] represented the model as in figure 1.

However, Bromberger [22] and Salmon [23,24] pointed out several very important shortcomings of this model of explanation, *viz*. its failure to account for explanatory relevance,[7] its permissive definition of laws of nature,[8] and its failure to account for explanatory asymmetries. Even though the problems of explanatory relevance and permissive definition of laws of nature are unique to the DN model, the problem of explanatory asymmetries puts any theory of explanation to the test. To illustrate the importance of explanatory asymmetries, consider one of the most famous objections to the DN model, i.e. the flagpole example, which is adapted from Bromberger [22]. Suppose we want to explain why the shadow of a flagpole is of certain length. According to the DN model, the *explanans* will include the antecedent conditions such as the height of the flagpole and the sun's elevation at a particular time, and together with the laws of optics, we would be able to explain why the flagpole shadow has a certain length, as is shown in figure 2.

But we can reverse this calculation and deduce the length of the flagpole or the elevation of the Sun based on the length of the shadow. The resulting 'explanation' would not strike us as particularly good, even though the argument is formally equivalent to the original one and the premises are empirically true. The trouble is, we might say, that the height of the flagpole and the position of the Sun do not depend on the length of the shadow. This has become known as the asymmetry problem in theories of explanation. A good theory of explanation should account for explanatory asymmetry.[9]

One obvious way to solve the asymmetry problem is to switch focus from inferential patterns to patterns of causation,

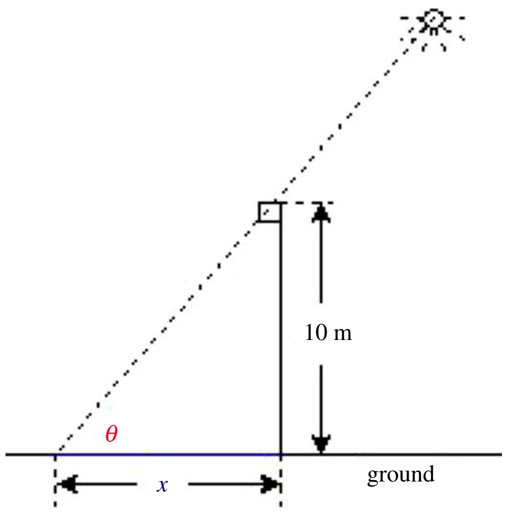

Figure 2. The flagpole example. (Online version in colour.)

or at the very least this is the lesson that many philosophers took from the flagpole problem. The general idea is that causes explain their effects and not the other way around, hence the causal patterns provide directionality to explanations, and thereby prevent explanatory symmetries. In the above flagpole example, this would imply that the flagpole's height causes the shadow's length, but not vice versa. This kind of asymmetry will have to be preserved across all the counterfactuals related to that explanation. This points toward another lesson one could take as well: it is the role of counterfactual dependence in explanations. If we take these lessons, a clearer path emerges when it comes to the solution of the asymmetry problem in topological explanations, namely, the counterfactual dependence is a structural feature of explanations that is available in both causal and non-causal explanations. Given that many philosophers argue that topological explanations are non-causal, because their *explanans* does not cite any causes, but rather mathematical properties of network topology [1,26–32], then appealing to causation or causal facts is not an option for solving the asymmetry problem in topological explanations. A better way to approach the problem is to inquire whether the counterfactual dependence in topological explanations can account for explanatory asymmetry. In his recent paper, Lange [33] argues that in several available counterfactual accounts of non-causal explanations, including the one espoused by Jansson [25], the counterfactual dependence alone is not sufficient to account for explanatory asymmetry. Jansson's contribution to this theme issue [9], takes up Lange's challenge directly, by focusing on the conditions of application in non-causal explanations. In the next section, two more possible bases of explanatory asymmetry in topological explanations are discussed; however, responding directly to Lange's [33] challenge is beyond the scope of this paper.

4. The bases of topological explanatory asymmetries

As we have seen, one of the most influential ways to solve the asymmetry problem for the DN model of explanation was to switch focus from inferential patterns to patterns of causation. The idea is that causes explain their effects, and not the other way around; and thus, in causal explanations the explanatory asymmetry is rooted in the directionality of causation. But it is not initially obvious what could be a basis of explanatory asymmetry in topological explanations, for it is unclear if they appeal to causation.

In this section, I provide several possible strategies to account for explanatory asymmetries in topological explanations, i.e. in terms of property, counterfactual and perspectival asymmetries, instead of asymmetry of causation. All three of these bases of topological explanatory asymmetries are tightly interconnected and rooted in the three conditions of the general theory of topological explanations.

One could conceive of property asymmetry in the following way. Suppose A explains B topologically. Then A is a topological property and B is a physical property. Consequently, the symmetry problem only arises if B topologically explains A. However, if B topologically explains A, then B must be a topological property as well. Some physical properties are topological properties thus whenever B is not a topological property the asymmetry holds. Recall from the example above about topological explanation of cognitive controllability. Cognitive controllability is not a topological property; therefore, the topological explanations of it exhibit property asymmetry.

The counterfactual asymmetry can be best understood in the following terms. Suppose A explains B topologically. Then B counterfactually depends on A. If there is an asymmetry problem, then A also counterfactually depends on B. But there will be cases in which B counterfactually depends on A but not *vice versa*. And in these cases, there will be no asymmetry problem.

The perspectival asymmetry has to do with the fact that although A is a relevant answer to an explanation-seeking question about B, sometimes B will not be an answer to the relevant explanation-seeking question Q about A, meaning that in the vertical mode, the fact that the system is stable will not be an answer to the question why the system is a small-world network. Equally, in the horizontal mode, the fact that the cognitive control has certain energy requirements does not answer why the local communicability measure has a certain value. Thus, reversing the direction of explanation would deem it non-explanatory, and the symmetry problem will be blocked.

Moreover, in the horizontal mode, explanatory perspectivism tends to figure prominently, as many 'local asymmetries' figure prominently. Very roughly, explanation-seeking questions are relevant, in part, because they serve scientists' interests. Often, the properties that interest scientists have asymmetries that are highly specific to the phenomena they are studying. As an illustration, consider cases in which various local features of brain networks determine certain behaviours or processes, e.g. asymmetry of information flow [34], or asymmetry of topological organization of the auditory cortex [35].

The asymmetry of information flow can be thought of in terms of the efficiency of network communication. Efficiency in this context is measured by shortest path lengths (the fewer edges that have to be traversed between two nodes the more efficient the communication between them), or in terms of strongest and more reliable connections between the nodes. For example, in complex systems such as the brain, communication between the brain region i and the brain region j is asymmetric if it can be achieved more efficiently from i to j, than the other way around. The explanation-seeking question here is: *Why are some propagation strategies of signalling pathways in communication networks better than others?* An answer to the question appeals to the graph-theoretical efficiency of communication between the nodes in a network ([34], p. 2).

Such an asymmetry is determined by using several network communication measures, e.g. navigation efficiency, diffusion efficiency and search information [34]. If region i (the source region) and region j (the target region) are not directly connected, the information flow from i to j must use one or more intermediate nodes. This implies that there could be different signalling pathways in the network, in virtue of which the communication efficiency may differ for each set of pathways. And the communication efficiency of a particular pathway may depend on the direction of the information flow. Seguin *et al*. conclude that these results 'may be primarily driven by specific properties of brain networks, rather than by aspects particular to one network communication measure' ([34], p. 10). Furthermore, they also suggest that 'complex organisational properties of nervous systems are necessary to shape the directionality of neural signaling' ([34], p. 10). This approach is particularly useful when applied to undirected topology of brain networks and in predicting large-scale neural signalling.

The direction of an explanation in this case follows the efficiency of network communication, which is best captured through the property of 'send–receive communication asymmetry' ([34], p. 1). Given that the explanation-seeking question was: *Why are some propagation strategies of signalling pathways in communication networks better than the other?*, The explanation is: because it is more efficient to send and receive signal from the region i to region j than the other way around, where the efficiency of direct and indirect send–receive communication is measured by the network navigation efficiency, diffusion efficiency and search information ([34], p. 2). In other words, the direction of explanation is the direction of send–receive efficiency—which only has one direction, since by the network communication efficiency definition, a node cannot simultaneously be both a sender and a receiver.

So, in this case, scientists are interested in efficient information flows. However, this is very much indexed to the specific question being asked, and hence is a 'local asymmetry' that is captured by the perspectival criterion of topological explanations.

It should be noted that the actual wiring patterns of course vary with particular complex systems; however, the explanation for why some signalling patterns are more efficient than other is based on underlying facts about the counterfactual dependence between the topological properties (such as network navigation efficiency, diffusion efficiency and search information) and the network dynamics. The explanation thus does not appeal to contingent facts about wiring in any particular system, but to counterfactual dependency between topological properties that determine network communication, and thus that holds independently from any particular system. The facticity criterion, or the fact that the system instantiates a certain topology, only tells us why such an explanation is true of a particular system, but the explanatory force stems from the counterfactual dependence between the topological properties (communication efficiency) and network dynamics (direction of neural signalling), and that is why the instantiation of topology in the system is not explanatory in itself [31].

A similar idea about the basis of topological directionality comes from consideration of asymmetric network embedding in auditory cortices. Structural and functional differences in auditory cortex affect the performance of a variety of sensory and cognitive tasks, such as speech and tonal processing ([35], p. 2656). The structural and functional lateralization of the auditory cortex are explained by appealing to topological properties that determine the network embedding in the auditory cortex. The explanation appeals to the property of topological centrality of the brain auditory networks, which is determined by using network communication measures—such as closeness centrality or nodal efficiency, which measure direct and indirect communication pathways between two hemispheres. The hemisphere with higher closeness centrality or nodal efficiency is better integrated into the overall network communication, which allows better global communication in the network, and most importantly this topological asymmetry is driven by differences in communication pathways between two hemispheres, i.e. the hemisphere that has higher nodal efficiency is better integrated into the global communication network, thus the difference in topological efficiency drives the functional lateralization in the cortex ([35], p. 2660). Here the explanation has the direction of the counterfactual dependence as well as a local asymmetry, i.e. asymmetric embedding of networks in auditory cortex determine the communication efficiency in both direct and indirect connections, and not the other way around. Equally, in terms of explanatory perspectivism, asymmetric embedding of networks in auditory cortex is the answer to the question of why there is a lateralization in cognitive function and not the other way around.

These are just two examples of the general idea about how counterfactual dependence and explanation-seeking questions provide bases of explanatory asymmetries in topological explanations, without appealing to causation.

Property, counterfactual and perspectival asymmetries dampen possible objections such as that topological explanations in the vertical mode will require some non-negligible ontic commitments in order to be considered asymmetric, e.g. that in some cases it might allow retrodiction which, in turn, requires causal assumptions.[10] This kind of objection can be preempted if one looks into the second criterion in the general theory of topological explanation, which stipulates that *A* topologically explains *B* in the vertical mode only if *A* describes a global topological property of the network, such that, had *A* had not obtained some general physical property *B* would not have obtained either. Thus, only the counterfactuals of this form should be considered (vertical) topological explanations, and the patterns that do not describe this exact type of counterfactual dependency should either not be considered cases of vertical topological explanations or are not explanations at all. As for the retrodiction, topological explanations in the vertical mode are synchronous, i.e. they rarely explain the dynamics, but rather some general empirical property of the system. To that effect, retrodiction in the vertical mode would require a different *explanandum*, which is a consequence of the explanatory perspectivism condition, i.e. that the kind of explanatory questions determine whether we use vertical or horizontal explanatory mode, in which case the explanation may as well be causal.

Having discussed some ways in which the general theory of topological explanations accounts for explanatory asymmetries in both vertical and horizontal modes, in the next section, I argue how the general theory and its accounting for the explanatory asymmetries can help to unify essential concepts of biological networks.

5. Generalizability of the epistemic norms for topological explanatory asymmetries

I have argued in this paper that determining whether a topological explanation is successful (or not) and understanding why it is successful (or not) can only be assessed through a general theory of topological explanations, which tells us what it is for *A* to topologically explain *B*. In §2, I laid out such a theory, according to which *A* topologically explains *B* if and only if:

1. (*Facticity*): *A* and *B* are approximately true; and
2. Either
   (a) (*Vertical mode*): *A* describes a global topology of the network, *B* describes some general physical property, and had *A* had not obtained, then *B* would not have obtained either; or
   (b) (*Horizontal mode*): *A* describes a set of local topological properties, *B* describes a set of local physical properties, and had the values of *A* been different, then the values of *B* would have been different.
3. (*Explanatory perspectivism*): *A* is an answer to the relevant explanation-seeking question Q about *B*, such that the Q determines whether to use vertical or horizontal explanatory mode.

This theory provides three criteria for evaluating the success of any topological explanation:

(a) The criterion concerning the facticity of topological explanations, i.e. what makes it true of a particular system.
(b) The criterion that governs two explanatory modes of topological explanations, i.e. the vertical and the horizontal;
(c) Finally, the third criterion tells us when to use the vertical and when to use horizontal mode. This last criterion is based on the idea of explanatory perspectivism, according to which the explanation-seeking questions determine whether we use vertical or horizontal mode in describing counterfactual dependencies.

One of the most important features of any theory of explanation is that it can account for explanatory asymmetries. In §3, I provided a background of the asymmetry problem in order to demonstrate why it is of foundational importance that a theory of explanation can account for explanatory asymmetries. In §4, I argued how the general theory of topological explanations accounts for explanatory asymmetries under two explanatory modes, by using examples of cognitive controllability and the asymmetry of information flow

[34], and the asymmetry of topological organization of the auditory cortex [35].

The final piece of my argument in this paper is to show how the general theory of topological explanations helps to unify essential concepts of biological networks. The best way to do so is by considering universal applicability of the three criteria that the theory provides.

The facticity criterion as it was shown works in both vertical and horizontal modes of topological explanations. Given that topological explanations are non-causal and that counterfactual dependency relations under horizontal and vertical modes hold independently from the contingent facts about any particular system, the facticity criterion tells us how this universally applicable explanation is to be successfully used in any particular case, regardless of the area of science.

The criterion about vertical and horizontal modes of topological explanations can be understood also in terms of counterfactual dependencies within-the-scale (horizontal) and between-the-scales (vertical), intra-modular (horizontal) and hierarchical-modular (vertical), or in general in terms of the distinction between local (horizontal) and global (vertical) network levels. The distinctions between different scales, modules and hierarchies of modules or between local and global levels are not limited to any particular class of cases or to any particular area of biology. Given this, they should be applicable to all networks, regardless of the discipline, and *ipso facto*, the criteria about vertical and horizontal modes of topological explanations are applicable across biological sciences.

Finally, the criterion about explanatory perspectivism also helps to delimit distinctively topological explanations from the other types of explanation or ensembles of explanations [36,37]. For example, sometimes we want to understand very particular phenomena (such as why certain connections exist) or perhaps very complex phenomena (such as development of certain dynamical constraints, e.g. the particular wiring costs). In those cases, we will have a combination of different kinds of explanation, e.g. mechanistic, dynamical, statistical and topological. This is particularly evident in the case of homophily, where regions with similar connectivity properties connect directly among themselves. We might ask why do they connect among themselves, and the answer is: owing to various environmental mechanisms in the evolution of a particular network. For example, an explanation of why particular nodes in a network are connected among themselves will appeal to local wiring costs and available energy. An explanation will also appeal to regularity rather than some teleological principle, e.g. connections that are established on a more regular basis will have lower wiring costs than connections that are established only occasionally. In this case, the explanation will not be distinctively topological, but rather mechanistic, despite involving some network concepts. More generally, when the explanation-seeking question is about how and why the system arrived at having certain topological properties, the explanation will not be distinctively topological, it will more likely be mechanistic. On the other hand, if the explanation-seeking question is why the system displays certain behaviours or dynamical properties given the topology that it instantiates, the explanation will be distinctively topological and it will employ either vertical or horizontal mode of describing counterfactual dependencies.

The examples that I used to illustrate these points come from computational neuroscience, biology and ecology, which indicate that these criteria are applicable to network explanations in all of those areas. More importantly, because these criteria are not limited to particular classes of cases or specific areas of biological sciences, they are in principle universally applicable across any sciences that use network explanations, thus their universal applicability may help to unify network concepts across biological sciences.

Data accessibility. This article has no additional data.

Competing interests. I declare I have no competing interests.

Funding. This research is funded by the European Commission, Horizon 2020 Framework Programme, Excellent Science, Marie Skłodowska-Curie Actions under REA grant agreement no 703662; project: Philosophical Foundations of Topological Explanations (Proposal acronym: TOPEX) and the project Neuroessentialisme: explication neuroscientifique et diffusion des neurosciences, le cas du neurodroit, funded by the Conseil régional Nouvelle-Aquitaine and Bordeaux Montaigne University.

Acknowledgements. I am grateful to Claus Hilgetag, Kareem Khalifa, Angela Potochnik, Thomas Polger and Cédric Brun for tremendously helpful comments and discussions of the ideas and various versions of this paper.

Endnotes

[1]A counterfactual is a statement describing a hypothetically different situation relative the actual state of affairs. In currently fashionable terms, it describes what-if-things-have-been-different relative to a postulated explanation.

[2]Technical term denoting the part of an explanation with which we are explaining (plural is *explanantia*).

[3]Technical term denoting the part of an explanation that describes what is being explained (plural is *explananda*).

[4]Topological explanations raise ontological questions about mathematical objects. Philosophers of mathematics cast this in terms of the so-called indispensability arguments, according to which we ought to rationally believe that mathematical entities that are indispensable in explaining physical facts are real in the same sense as other unobservable theoretically postulated entities such as quarks, dark matter or black holes [3–8]. However, this paper shall remain neutral about this issue, because addressing it exceeds its scope.

[5]The idea of using network control theory in this context comes from control theory in engineering. Gu *et al.* explicitly acknowledge it: 'Network control theory is a branch of traditional control theory in engineering that addresses the question of how to control a system whose components are linked in a web of interconnections; here the term control indicates perturbing a system to reach a desired state'. ([12], p. 8).

[6]The measure of network communicability assesses to what extent the nodes in a network are connected indirectly. Network communicability is calculated by quantifying weighted sum of walks of all lengths in the communicability matrix, with a direct implication that longer lengths require more energy, and the shorter ones require less energy. Now we see that the topology that is described by communicability measures can affect energy requirements for changing the trajectories in the state space, and therefore how from particular values of these measures it is possible to find both the exact energy requirements and also the minimal ones.

[7]According to the explanatory relevance objection, the DN fails to distinguish between relevant and irrelevant antecedent facts, thus the following case would be considered a successful explanation: (1) birth control pills prevent pregnancy, (2) Jones (a male) has taken birth control pills, (3) therefore, Jones did not get pregnant. Clearly, Jones did not get pregnant because he does not have appropriate reproductive organs to become pregnant in the first place, thus the DN model focuses solely on inferential patterns, without providing a norm for explanatory relevance [24].

[8]Namely, the definition of the laws of nature does not distinguish between accidental generalizations and lawful generalizations.
[9]However, some phenomena seem to permit symmetric explanations, e.g. in one system of gas, decrease of volume can account for increased pressure; in another, increased pressure can account for decreased volume; or some laws of nature, especially in physics also seem to provide symmetric explanations [25]. A good theory of explanation should account for such cases as well. As far as I can tell this does not apply to topological explanations (in either vertical or horizontal mode).
[10]I thank Referee 1 for raising this point.

References

1. Kostić D. 2019 Minimal structure explanations, scientific understanding and explanatory depth. *Perspect. Sci.* **27**, 48–67. (doi:10.1162/posc_a_00299)
2. van den Heuvel MP, Sporns O. 2013 Network hubs in the human brain. *Trends Cogn. Sci.* **17**, 683–696. (doi:10.1016/j.tics.2013.09.012)
3. Baker A. 2005 Are there genuine mathematical explanations of physical phenomena? *Mind* **114**, 223–238. (doi:10.1093/mind/fzi223)
4. Field HH. 1989 *Realism, mathematics and modality*. Oxford, UK: Blackwell.
5. Lyon A. 2012 Mathematical explanations of empirical facts, and mathematical realism. *Australas. J. Philos.* **90**, 559–578. (doi:10.1080/00048402.2011.596216)
6. Saatsi J. 2011 The enhanced indispensability argument: representational versus explanatory role for mathematics in science. *Br. J. Philos. Sci.* **63**, 143–154. (doi:10.1093/bjps/axq029)
7. Saatsi J. 2016 On the 'indispensable explanatory role' of mathematics. *Mind* **125**, 1045–1070. (doi:10.1093/mind/fzv175)
8. Psillos S. 2010 Scientific realism: Between platonism and nominalism. *Philos. Sci.* **77**, 947–958. (doi:10.1086/656825)
9. Jansson L. 2020 Network explanations and explanatory directionality. *Phil. Trans. R. Soc. B* **375**, 20190318. (doi:10.1098/rstb.2019.0318)
10. Watts DJ, Strogatz SH. 1998 Collective dynamics of 'small-world' networks. *Nature* **393**, 440. (doi:10.1038/30918)
11. Kaiser M, Hilgetag CC. 2004 Edge vulnerability in neural and metabolic networks. *Biol. Cybern* **90**, 311–317. (doi:10.1007/s00422-004-0479-1)
12. Gu S *et al.* 2015 Controllability of structural brain networks. *Nat. Commun.* **6**, 8414. (doi:10.1038/ncomms9414)
13. Betzel RF, Gu S, Medaglia JD, Pasqualetti F, Bassett DS. 2016 Optimally controlling the human connectome: the role of network topology. *Sci. Rep.* **6**, 30770. (doi:10.1038/srep30770)
14. Medaglia JD, Zurn P, Armstrong WS, Bassett DS. 2016 Mind control: frontiers in guiding the mind. http://www.arxiv.org>qbio>arXiv:1610.04134.
15. Gu S *et al.* 2015 Controllability of structural brain networks. *Nature Commun.* **6**, 8414. (doi:10.1038/ncomms9414)
16. Achinstein P. 1977 What is an explanation? *Amer. Philos. Quart.* **14**, 1–15.
17. van Fraassen BC. 1980 *The scientific image*. Oxford, UK: Oxford University Press.
18. Kostić D. 2017 Explanatory perspectivalism: limiting the scope of the hard problem of consciousness. *Topoi* **36**, 119–125. (doi:10.1007/s11245-014-9262-7)
19. Hilgetag CC, Goulas A. 2016 Is the brain really a small-world network? *Brain Struct. Funct.* **221**, 2361–2366. (doi:10.1007/s00429-015-1035-6)
20. Hempel CG, Oppenheim P. 1948 Studies in the logic of explanation. *Philos. Sci.* **15**, 135–175. (doi:10.1086/286983)
21. Hempel C. 1965 *Aspects of scientific explanation and other essays in the philosophy of science*. New York, NY: Free Press.
22. Bromberger S. 1966 'Why questions.' In *Mind and cosmos: essays in contemporary science and philosophy* (ed. R Colodny), pp. 86–111. Pittsburgh, PA: University of Pittsburgh Press.
23. Salmon WC. 1971 *Statistical explanation and statistical relevance* (*Vol. 69*). Pittsburgh, PA: University of Pittsburgh Press.
24. Salmon WC. 2006 *Four decades of scientific explanation*. Pittsburgh, PA: University of Pittsburgh Press.
25. Jansson L. 2015 Explanatory asymmetries: laws of nature rehabilitated. *J. Philos.* **112**, 577–599. (doi:10.5840/jphil20151121138)
26. Jones N. 2014 Bowtie structures, pathway diagrams, and topological explanation. *Erkenntnis* **79**, 1135–1155. (doi:10.1007/s10670-014-9598-9)
27. Huneman P. 2010 Topological explanations and robustness in biological sciences. *Synthese* **177**, 213–245. (doi:10.1007/s11229-010-9842-z)
28. Huneman P. 2018 Outlines of a theory of structural explanations. *Philos. Stud.* **175**, 665–702. (doi:10.1007/s11098-017-0887-4)
29. Rathkopf C. 2018 Network representation and complex systems. *Synthese* **195**, 55–78. (doi:10.1007/s11229-015-0726-0)
30. Darrason M. 2018 Mechanistic and topological explanations in medicine: the case of medical genetics and network medicine. *Synthese* **195**, 147–173. (doi:10.1007/s11229-015-0983-y)
31. Kostic D. 2018*a* The topological realization. *Synthese* **195**, 79–98. (doi:10.1007/s11229-016-1248-0)
32. Kostic D. 2018*b* Mechanistic and topological explanations: an introduction. *Synthese* **195**, 1–10. (doi:10.1007/s11229-016-1257-z)
33. Lange M. 2019 Asymmetry as a challenge to counterfactual accounts of non-causal explanation. *Synthese*. (doi:10.1007/s11229-019-02317-3)
34. Seguin C, Razi A, Zalesky A. 2019 Send–receive communication asymmetry in brain networks: inferring directionality of neural signalling from undirected structural connectomes. *BioRxiv*, 573071.
35. Mišić B *et al.* 2018 Network-based asymmetry of the human auditory system. *Cereb. Cortex* **28**, 2655–2664. (doi:10.1093/cercor/bhy101)
36. Kauffman SA. 1993 *The origins of order: self-organization and selection in evolution*. New York, NY: Oxford University Press.
37. Rivelli L. 2019 Multilevel ensemble explanations: a case from theoretical biology. *Perspect. Sci.* **27**, 88–116. (doi:10.1162/posc_a_00301)