ACCEPTED IN SYNTHESE
PRE-PRINT-PLEASE CITE THE PUBLISHED VERSION

Authors: Daniel Kostić, PhD and Willem Halffman, PhD.

Title: Mapping Explanatory Language in Neuroscience.

Affiliations:
Daniel Kostić
Institute of Philosophy
University of Leiden
Nonnensteeg 1-3
2300 RA Leiden
The Netherlands

Willem Halffman
Institute for Science in Society (ISiS)
Radboud University
P.O. Box 9010
6500 GL, Nijmegen
The Netherlands

E-mails: daniel.kostic@gmail.com
w.halffman@science.ru.nl

Phone: +33 (0) 768890295.

Conflict of Interest Statement

We declare that we have no conflict of interest.

Acknowledgements: We would like to thank Guillaume Cabanac and Wendeline Swart for preparing the data from the Dimensions, and to Guillaume Levrier for the help with graphics. We are grateful for very insightful comments on the manuscript to Charles Pence and Samuel Fletcher. We thank Digital Science for making Dimensions data (https://dimensions.ai) available for research.

Funding: Daniel Kostić would like to acknowledge funding by the Radboud Excellence Initiative.

Mapping Explanatory Language in Neuroscience

Abstract

The philosophical literature on scientific explanation in neuroscience has been dominated by the idea of mechanisms. The mechanist philosophers often claim that neuroscience is in the business of finding mechanisms. This view has been challenged in

numerous ways by showing that there are other successful and widespread explanatory strategies in neuroscience. However, the empirical evidence for all these claims was hitherto lacking. Empirical evidence about the pervasiveness and uses of various explanatory strategies in neuroscience is particularly needed because examples and case studies that are used to illustrate philosophical claims so far tend to be hand-picked. The risk of confirmation bias is therefore considerable: when looking for white swans, all one finds is that swans are white. The more systematic quantitative and qualitative bibliometric study of a large body of relevant literature that we present in this paper can put such claims into perspective. Using bibliometric tools, we identify the typical linguistic patterns used in the alleged mechanistic, dynamical, and topological explanations in the literature, their preponderance and how they change over time. Our findings show abundant use of mechanistic language, but also the presence of a significant neuroscience literature using topological and dynamical explanatory language, which grows over time and increasingly differentiates from each other and from mechanistic explanations.

1. Introduction

The philosophy of science has been marked by an ever-growing interest in scientific explanations. This interest is especially unsurprising in the philosophy of neuroscience, given the sheer diversity of modelling and explanatory practices in neuroscience (Gold and Roskies 2008). The philosophical literature on scientific explanation in neuroscience has been dominated by the idea of mechanisms (Craver 2007; Bechtel and Richardson 2010; Glennan 2017). The basic idea can best be captured by the following definition of a minimal mechanism (Glennan 2017, 17):

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

The mechanist philosophers often claim that all explanations in neuroscience are ultimately mechanistic in the above sense, or, at the very least, that they conform to various degrees of completeness of this definition, e.g., there could be full-fledged mechanisms, partial mechanisms or mechanistic sketches (Piccinini and Craver 2011). Furthermore, anything that does not fit this definition, or a degree of completeness thereof, is not considered an explanation at all (Craver 2016). Other diverse explanatory strategies are thereby reduced to a single mechanist formula and hence we call this set of claims "explanatory imperialism" (Kostić 2022).

Mechanistic imperialism has been challenged by various arguments which show that there are scientific explanations in science in general, and in neuroscience in particular, that do not conform to the mechanistic mould. Among the contenders that generated the most philosophical literature are the dynamical (Weiskopf 2011; Chemero and Silberstein 2008; Vernazzani 2019; Favela 2020; 2021; Stepp, Chemero, and Turvey 2011; Gervais 2015; Verdejo 2015; Venturelli 2016) and topological explanations (Kostić 2022, 2020a, 2020b, 2018; Kostić and Khalifa 2021, 2022; Ross 2021).

We acknowledge that there are many other non-mechanistic kinds of explanations across the sciences, e.g., computational (Chirimuuta 2014), statistical (Walsh 2014; Walsh, Lewens, and Ariew 2002), interventionist (Woodward and Hitchcock 2003; Hitchcock and Woodward 2003), mathematical and in general non-causal explanations (Lange 2013), minimal model and optimality explanations (Batterman 2010; Batterman and Rice 2014; Rice 2021), and many other. However, here we focus on dynamical and topological explanations for three reasons: 1) they directly and in depth challenge mechanistic imperialism, especially in neuroscience; 2) these explanations use a relatively distinct repertoire to express explanatory relations, and such

repertoire can be traced in the language used in scientific literature, and finally, 3) our aim in this paper is not to represent the full range of explanatory repertoires in the neurosciences, but to demonstrate that important competitors for mechanist explanations exist and thrive in scientific practice.

Mechanistic imperialism can be interpreted in two ways. The first is that explanations may look non-mechanistic, but these can, in principle, always be interpreted as mechanistic by using the epistemic-normative frameworks developed by the new mechanists. In this view, the scientists may use non-mechanistic terms in describing their explanatory practices, but these descriptions actually conform to the mechanists' conception of scientific explanation. The second is a more empirical claim that the repertoire of mechanistic explanations prevails in neuroscientific practice. In the former interpretation, the pervasiveness of one or the other scientific explanations is determined solely through conceptual analysis, whereas in the latter, it requires empirical evidence. As this paper investigates explanatory repertoires empirically, i.e., in the language of research papers, it addresses directly the latter, empirical issue.

The empirical claims about the pervasiveness of one or the other kind of explanations in neuroscience require empirical evidence, which so far has not been forthcoming. The importance of empirical evidence about pervasiveness and uses of "mechanisms" or any other kind of explanation in neuroscience is particularly needed because examples and case studies that are used to illustrate philosophers' claims do not represent a statistically relevant sample, even if taken all together. Since demonstrations of the pervasiveness of different kinds of explanation in the philosophical literature rely on handpicked examples, the risk of confirmation bias is considerable: when looking for white swans, all one finds is that swans are white. The more systematic quantitative and qualitative bibliometric study of a large body of relevant literature that we present in this paper can put such claims into perspective by investigating:

1) What are typical mechanistic, dynamical, and topological expressions used in neuroscience papers?
2) What is the preponderance of mechanistic, dynamical, and topological explanations in the neuroscience literature?
3) How does the preponderance of these explanatory patterns in neuroscience change over time?

In this study, we first defined strings of words to identify explanatory language patterns in a qualitative analysis, and then searched for these strings in a large neuroscience corpus from the Dimensions.ai repository. In a second step, we analysed the distribution of typical language patterns in the corpus to provide comprehensive and empirically grounded insights into the explanatory landscape of neuroscience.

In order to provide a philosophical context for our study, in the next section we characterize more precisely each of the three kinds of explanations. In the interest of space, we skip an overview of the debates between the mechanistic imperialist and proponents of dynamical and topological explanations because review literature of these debates is abundant (Kostić, Hilgetag, and Tittgemeyer 2020; Kostić 2018a, 2019, 2022; Khalifa et al. 2022).

2. Mechanistic, dynamical and topological explanations

2.1.Mechanistic explanation

According to some of the most prominent mechanist philosophers "Biologists seek mechanisms that produce, underlie, or maintain a phenomenon" (Craver and Darden 2013, 72)

The most influential definition of mechanistic explanation comes from an early paper by Machamer and colleagues (Machamer, Darden, and Craver 2000, 3):

Mechanisms are entities and activities organized such that they are productive of regular changes from start or set-up to finish or termination conditions.

In this definition, entities in mechanisms could be neurons in a brain that are organized in a certain way, e.g., connected into neural populations that make up brain regions. But they also have to do something: they have to produce or change things through some activity. For example, neurons release neurotransmitters in order to propagate signals through neuronal assemblies. This is where the comparison with some everyday notions of mechanisms might be useful: a mechanical watch that has stopped ticking is not a mechanism in the above sense, because even though it has all the entities and components necessary for a mechanism, it lacks an activity. The activities that produce change in a mechanism are often linear in time, i.e., organized in sequences in which earlier stages produce later stages. They can also be cyclical, e.g., the Krebs cycle in the metabolism of sugar, in which some chemical compounds leave the mechanism at key junctures, but their residue is used at the next stage to continue the process. Finally, mechanisms can be described as underlying a phenomenon we want to explain. For example, the Hodgkin-Huxley model of action potential that explains the basic mechanism of signal propagation between neurons does not produce the phenomenon; it rather underlies it, or implements it (Craver and Darden 2013, 50). All these ideas can be best described by the so-called Craver diagram (Figure 1).

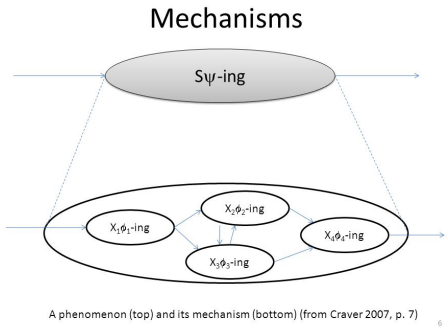

Figure 1: Craver diagram. Linear mechanisms at the bottom; a phenomenon at the top is constituted or implemented by the mechanisms at the bottom, which is represented with dotted lines between two levels.

For the purpose of this study, an exposition of more sophisticated distinctions of mechanisms would be superfluous. The most important lesson to take from this is that, typically, entities in a mechanism are linguistically described with nouns and activities with verbs. In neuroscience, these entities can be neurons or neuronal assemblies, causing phenomena by their activity. An example would be: cell membranes, ion channels, *Na* levels (i.e., *explanans* consisting of entities or components) produce, generate, or underly (or a verb expressing causation) action potentials (i.e., a higher-level *explanandum*).

2.2. Topological explanation

Topological explanations (proper) are a relatively recent development in the sciences that was enabled by a seminal paper by Watts and Strogatz (1998), and soon followed by several other key papers in different areas of science (Cupal, Kopp, and Stadler 2000; Barabasi and Albert 1999; Barabási and Oltvai 2004; Stadler and Stadler 2004).[1] Neuroscience did not lag behind, and the publication of a highly influential paper by Sporns and colleagues (Sporns, Tononi, and Kötter 2005) marked the birth of so-called network neuroscience and the origin of topological explanations in neuroscience. In the growing philosophical literature on topological explanations, there is only one account that provides necessary and sufficient conditions for a topological explanation in neuroscience (Kostić 2020a). According to this account:

a's being *F* topologically explains why *a is G* if and only if:
(T1) *a* is *F* (where *F* is a topological property)*;*
(T2) *a* is *G* (where *G* is a physical property)*;*
(T3) Had *a* been *F'* (rather than *F*), then *a* would have been *G'* (rather than G);
(T4) *a is F* is an answer to the question why is *a,* G?

Networks are collections of nodes and edges, and topological properties are their mathematically quantifiable patterns of connectivity. In this framework, the T1 and the T2 conditions simply mean that the same system can have both a physical and a topological property. For example, a brain which is denoted as an *a* in the scheme above can be both computationally efficient (i.e., it uses optimal amount of energy for processing information), which is its physical property *G*, and when represented as a network of anatomical connections it also is a small-world network, which is its topological property *F*. The T1 and T2 thus concern the representation of the system.

The T3 in Kostić's scheme describes a counterfactual dependence between a system's topological and physical properties. In the example with the brain, the T3 tells us that the brain would not have been computationally efficient if it had a random or a regular topology instead of the small-world topology that it actually has. The T3 condition hence concerns the explanation because it tells us *why something is the case*.

Finally, the fourth condition provides criteria for using the counterfactual. Such criteria are perspectival, in the sense that they provide a context which makes it intelligible why some empirical property *G* counterfactually depends on a network connectivity pattern, which is expressed as its topological property *F* (Kostić forthcoming). Relevant linguistic patterns in topological explanation therefore will be expressed as phrases containing nouns which denote topological properties and verbs denoting some form of a dependence. In the neuroscience literature, such an explanation would be expressed as a proposition in which a physical phenomenon (e.g., computational efficiency, robustness, or controllability) counterfactually depends on topological properties (e.g., a small-world, scale-free topology, or in general a connectivity pattern).

2.3.Dynamical explanation

A dynamical explanation is typically used to explain evolution of a chaotic system, or changes in a chaotic system over time. The possible states of a system are described as its state space, in which actual changes over time, from one state to another, form a trajectory.

[1] Even though some proto-topological explanatory language, based mainly on Euler and Erdős'work, was present in the literature for much longer.

By using differential equations of nonlinear dynamical systems theory, it is possible to quantify these changes over time, which uniquely determine the subsequent states of the system, e.g., in systems becoming synchronized. Since the dynamical explanation focuses on the mathematical properties of a dynamical model, entities, activities and microphysical causal details of underlying mechanisms are explanatorily idle (Chemero and Silberstein 2008; Favela 2020; 2021; Gervais 2015; Khalifa et al. 2022; Stepp, Chemero, and Turvey 2011; Venturelli 2016; Verdejo 2015). As such, dynamical explanation is typically used to explain the global behaviour of a system. For example, in neuroscience, a dynamical explanation is used to explain why bimanual coordination (synchronous wagging of the same fingers on both hands) is in, or out of phase. To that effect, a relevant linguistic pattern in dynamical explanations will be a noun denoting a dynamical property and a verb such as "to determine" or to "shape".

3. Methods and Data

In this section, we explain how we were able to detect these three different explanatory patterns in a large body of neuroscience literature. We used basic text mining tools to identify typical word patterns that resemble explanatory language. Our approach had two stages. In the first stage, we used three sets of twenty neuroscience papers each, which were cited as typical examples of mechanistic, dynamical and topological explanations, respectively, in the philosophical literature that discusses these three types of explanations (see appendices 1 and 2). These three sets were used as 'training sets' to identify word patterns presumably typical of each of these explanations, to be later tested in the larger corpus of neuroscience literature. We decided not to start with a top-down hypothetical list of word patterns that could be expected to express explanation according to three philosophical accounts of scientific explanation (Fletcher et al. 2021; Bonino et al. 2022; Malaterre, Chartier, and Pulizzotto 2019; Mizrahi and Dickinson 2022a; 2022b), in order to avoid possible interpretative bias. Instead, our approach was bottom-up, as we started with the actual explanatory language used in neuroscience papers.

The full text of the three training sets was uploaded to the free text mining application Voyant-tools.org. This web-based application provides easy tools for calculating word frequencies, word co-occurrence, or quick access to the context of particular words in the text (e.g., fifteen words before and after a word of interest, which is easily expandable to a larger context if necessary). These tools allowed us to identify meaningful terms and count recurring word patterns, excluding stop-words such as "the", "a", or "was", digits for page numbers or years of publication (1995, 2, 43, etc.), connectives such as "and", "or", bibliographical abbreviations such as "et al.", etc. Among the most frequently occurring words, we identified terms that seemed to refer to elements typical for dynamical, topological, or mechanistic explanations, i.e., *explanans*, *explanandum*, or explanatory relation terms between them.

To analyse the most frequent terms, we used stemmed words, e.g., in order to count "analysing", "analysis", "analytic" etc., as all belonging to the same term "analy*", in which the asterisk expresses an arbitrary suffix. A first joint inspection indicated that the non-random *explanantia* terms, unique to each of the three types of explanations, appeared among the twenty-five most frequent terms in each training set, since after the twenty-five most frequent terms we observed that the terms became paper-specific and no longer related to explanatory terms. This process revealed some terms that seemed to occur uniquely in one of our training sets, but also terms that occurred most frequently in the overlap, in one or both of the other sets. Although the philosophical literature presented the neuroscience articles in our training sets as typical of particular explanatory styles, these neuroscience articles do also frequently contain words that that are not pertinent to the analysis of explanatory language. Hence, other terms that occurred frequently in one of the training sets, just seemed accidental, such as "mouse" or

"visual*" (see Figure 2). Expressions typical of topological, dynamical, or mechanistic explanations can therefore not be easily derived from the mere frequency of particular words in such a training set.

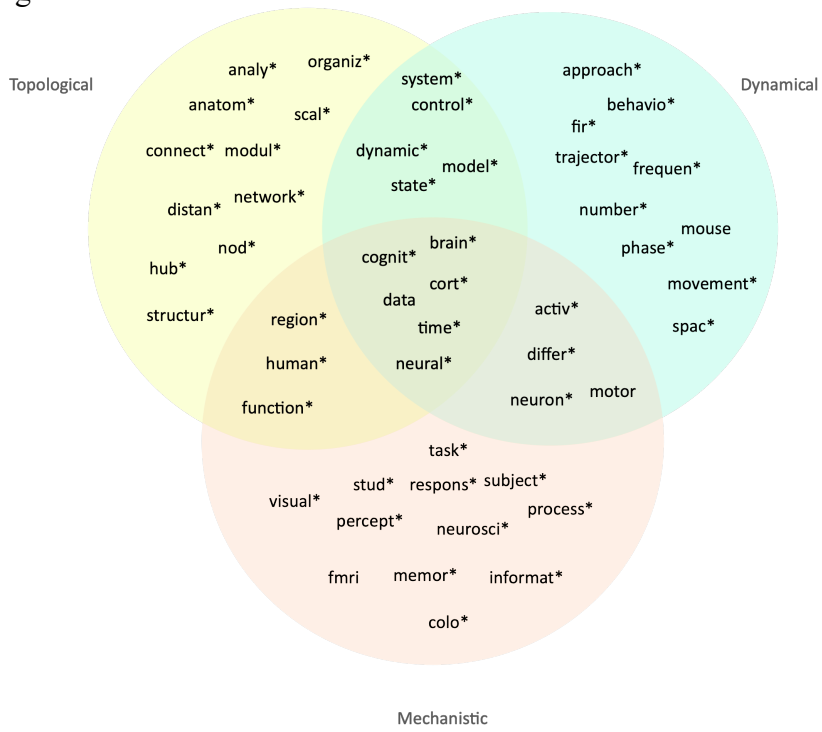

Figure 2. Most frequent 25 terms in the topological, dynamical, and mechanistic training sets and their overlaps between the sets.

After joint inspection of the terms in their contexts by both authors, it also became clear that the *explanandum* would not be distinctive for the type of explanation: all three types of explanations often aim for the same *explananda* terms, such as, for example, motor functions or cognition. Rather, our reading of the text in the training set showed that *explanantia* terms co-occurring with explanatory verbs express the most distinctive explanatory relations. For example, in the phrase "dynamics also create completely new behavioural constraints", the *explanans* "dynamics" and explanatory term "create" are typical of dynamical explanatory language, while "behavioural constraints" might conceivably also be explained in mechanistic or topological terms. In identifying typical expressions, we aimed for distinct word patterns, typical for specific explanatory schemes, rather than capturing all explanations.

With this tentative long-list of *explanantia* terms for each of the training sets, developed in our joint inspection, we individually set out to identify such typical word patterns, i.e., identifying phrases containing *explanantia* in combination with explanatory relations, most often verbs (see appendix 3). Our long list included *explanantia* terms such as "time", "non-linear", "state" for dynamic papers; "architecture", "topology", or "connectivity" for topological ones; "neuro", "neural", "activity" for mechanistic papers. After identifying phrases that we both independently judged as characteristic, we discussed each phrase until we reached a consensus about which phrases were characteristic examples of the three explanatory styles. For example, we removed explanatory terms that express a vague relation (e.g., "correlates with…", "is associated with…") without clear explanatory relationship. Some *explanantia* candidates, such as the term "time" or "non-linear", had to be removed because they returned too many phrases that were not explanatory, but referred to methodological or technical descriptions.

Our selection is thus based on our joint understanding of what constitutes mechanistic, topological and dynamical explanations, as specified in theoretical section above. If one of us

raised doubt about whether an explanation was, for example, truly topological, then the expression was removed. Although this admittedly involves human judgement, in this way we prioritised clear-cut expressions, at the expense of losing many less explicit ones. Although one of us has taken a position in the philosophical debate on explanations in previous work, the other author has no intellectual stake in these debates.

In the remaining phrases, we then counted the word distance between the *explanans* and the explanatory term, i.e., the number of words between the characteristic terms. For example, in "activity controls", the word distance between "activity" and "controls" is zero. In "activity that generally controls", the distance between "activity" and "controls" is two, i.e., two words, noted as "activity controls~2". We agreed to consider only single digit distances, because in principle, data noise and a possibility of false negatives increase with the higher limit on distance. The actual distances that we found in our training sets are listed in the appendix 3.

The word patterns were expressed as a complex search string with which to search the larger neuroscience literature via the Dimensions.ai database, which covers an exceptionally large number of research papers (almost 130 million). We limited the research to the period 1990-2021 and to papers labelled as "neurosciences" in the database (i.e., category 1109), totalling 2.199.526 papers. Since Dimensions is so comprehensive, we surmise that a very similar procedure for selecting a corpus can be used for just about all of the relevant literature (at least in English), avoiding random sampling errors. The delineation of research fields in bibliographic databases is generally somewhat ambiguous, but this should not fundamentally affect the results of our analysis. Apart from the size advantage, Dimensions allows for searches in the abstract and full text of articles, to the extent that Dimensions has access to them.

Three complex search strings were composed, one for each type of explanation, with all combinations of *explanantia* terms and explanatory relation terms unique for each type (for the search strings with the most hits in each corpus, see Table 1). The search added the word distance after each first term in the expression and combined all the expressions with a Boolean "OR" e.g., ("contribution of connectivity~2") OR ("depends on connectivity~1") OR … In other words: the database would return all articles that contain at least one of the word patterns that are typical of each of three different kinds of explanation identified in our training set. When ran through the large corpus of over two million papers in the Dimensions database, these search strings had found different number of papers for each kind of explanations. Table 1 represents the search strings and how many papers each search string retrieved from the Dimensions. A schematic of the method is provided in Figure 3.

Top 20 components	string (Mechanistic in bold)	No. of papers	string (Dynamical in bold)	No. of papers	string (Topological in bold)	No. of papers
1	**"activity modulate~5"**	69.489	"activity control~3"	8.283	**"connectivity predict~8"**	8.141
2	"dynamic lead to~2"	69.489	"connectivity confering~4"	8.283	"determined by dynamic~3"	8.141
3	"connectivity enhance~2"	69.489	**"dynamic influence~ 6"**	8.283	"Result of activity~1"	8.141

4	**"activity control~3"**	59.534	**"dynamic effect on~5"**	6.429	"activity Responsible for~4"	4.495
5	"connectivity confering~4"	59.534	"activity dictates~1"	6.429	**"connectivity Influence~5"**	4.495
6	"dynamic influence~6"	59.534	"connectivity constraining~4"	6.429	"governed by dynamic~2"	4.495
7	"topological predicting~8"	48.486	"connectivity constrain~4"	5.511	"neuron evoked~0"	2.947
8	**"neural underlying~2"**	48.486	"activity affects~1"	5.511	**"connectivity shape~6"**	2.947
9	"topological responsible for~5"	48.218	**"dynamic determine~6"**	5.511	"dynamical influence~6"	2.947
10	**"neural substrate~0"**	48.218	"connectivity confer~4"	4.184	"Generate by activity~4"	2.213
11	**"neural basis~0"**	47.701	**"dynamic predict~6"**	4.184	"underlying dynamic~1"	2.213
12	"topological shape~6"	47.701	"activity controlling~3"	4.184	**"connectivity predicting~8"**	2.213
13	"connectivity facilitate~2"	30.702	"connectivity determine~1"	3.133	"dynamical work to~1"	2.210
14	"dynamic work to~1"	30.702	**"dynamic explain~6"**	3.133	"neuron control~1"	2.210
15	**"activity results in~4"**	30.702	"activity regulate~1"	3.133	**"Role of connectivity~3"**	2.210
16	**"neural control~1"**	30.165	"generate by activity~4"	2.815	"depend on dynamic~2"	1.782
17	"topological predict~8"	30.165	**"underlying dynamic~1"**	2.815	"activity precedes~0"	1.782
18	"dynamical work to~1"	27.234	"connectivity predicting~8"	2.815	**"connectivity impact~4"**	1.782
19	**"neuron control~1"**	27.234	"connectivity determining~1"	2.098	"Depends on activity~3"	1.608
20	"role of connectivity~3"	27.234	"activity produce~4"	2.098	**"connectivity play role~6"**	1.608

Table 1.: The actual search strings and the number of papers each of them retrieved when ran through the Dimensions.ai database. The strings identified in total: 417.422 mechanistic papers, 34.655 dynamical papers, and 32.961 topological papers. We highlighted in bold the strings we consider typical for each corresponding explanatory repertoire, which indicates that the explanatory language is not mutually exclusive. The tildes with a number express the total word distance between the characteristic words for each specific search term.

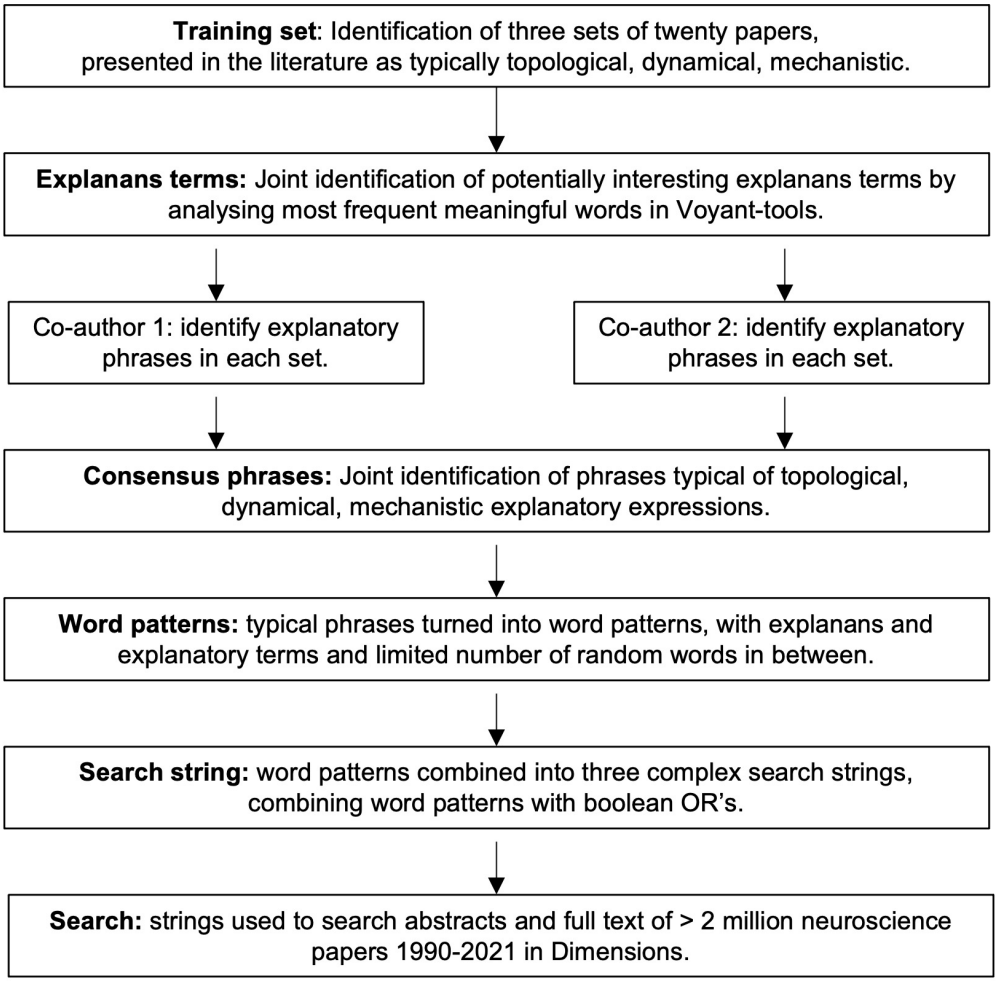

Figure 3. Method used to identify papers with typically topological, dynamical, or mechanical explanatory language.

4. The Results

The search in Dimensions returned a total of 443.966 papers, out of which 94% also had abstracts. Among the search results, the mechanistic set was by far the largest, while the dynamic and topological search strings each returned just over 30.000 results (see Table 2). This may be the result of overly specific search strings for these two latter sets, less specific mechanistic search strings, or an actual indication of the minority share of the dynamical and the topological explanations: we do not claim we have captured *all* papers of each explanatory type, just three characteristic sets. Our results should therefore not be read as an accurate representation of shares of either of these explanations in the literature, but as indicators of their presence and, as we show below, of their relative development over time.

	Number of papers	With abstracts
Total neuroscience papers	2.199.526	
Mechanistic search string	417.422	391.305

Dynamic search string	34.655	32.024
Topological search string	32.961	31.058
Total unique papers in our corpus	443.966	415.401

Table 2. The total number of neuroscience papers in Dimensions (1990-2021), and the number of papers identified through our search strings.

The actual number of papers matching our search strings for all three kinds of explanations per year since 1990 is shown in the Figure 4. These absolute numbers are misleading when we attempt to spot trends, as in this same period the total number of neuroscience papers also grew significantly.

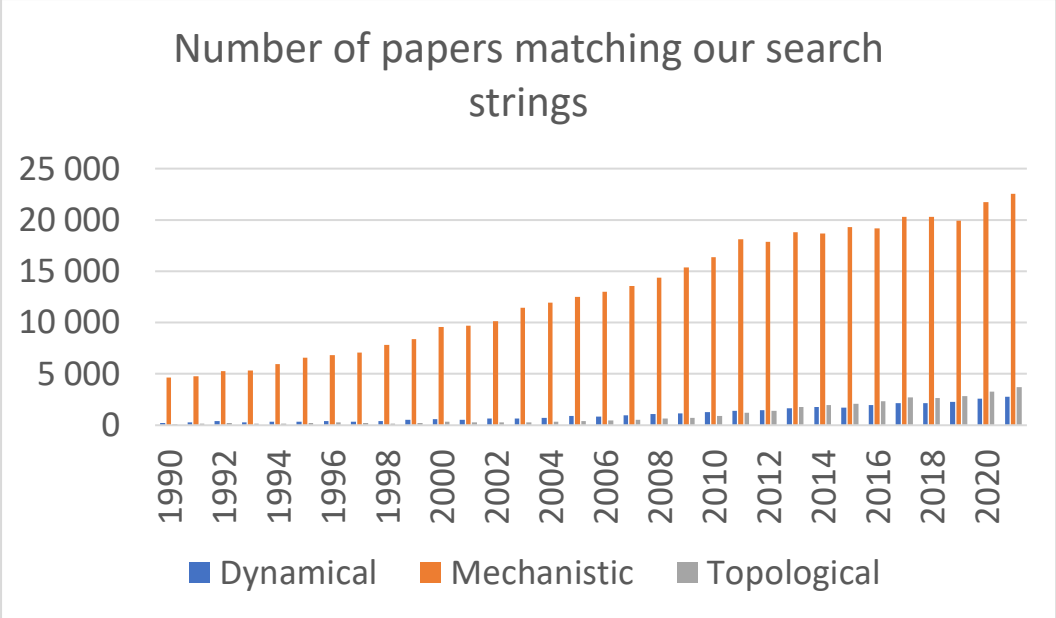

Figure 4. The total number of papers per year, from 1990 to 2020, matching our search strings.

The growing number of neuroscience papers in Dimensions since 1990, is shown in Figure 5.

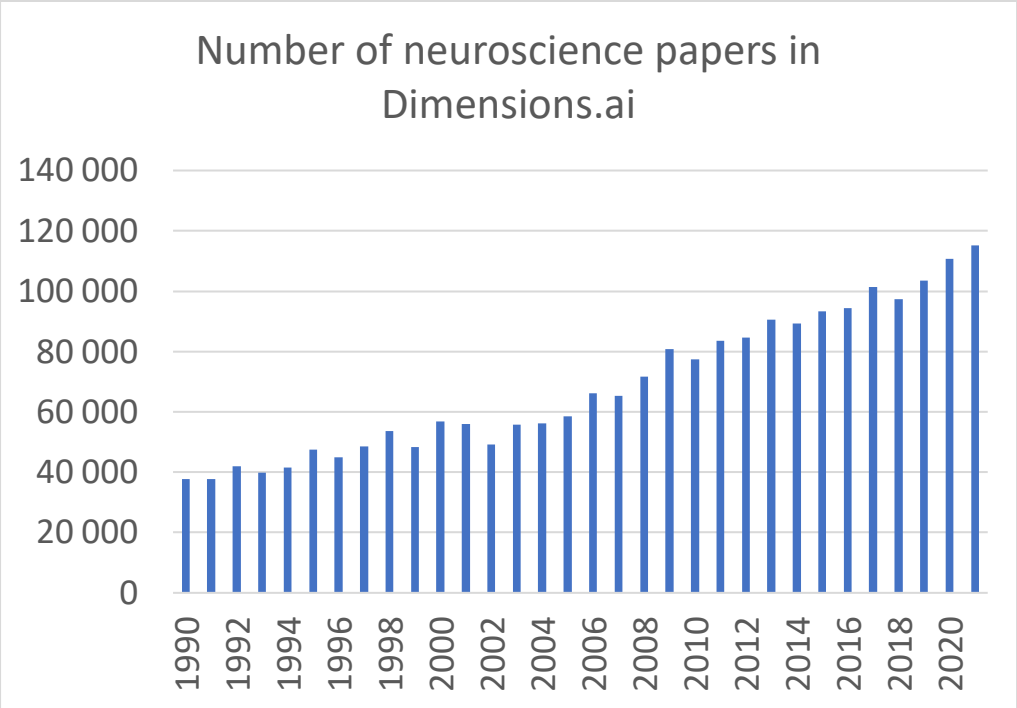

Figure 5. The total number of neuroscience papers from 1990 to 2020 in the Dimensions.ai.

As a more meaningful representation of how the three types of explanations develop over time, the share of mechanistic, dynamical and topological explanations in the total number of neurosciences papers is shown in the Figure 6. Once again, the actual share depends very much on the accuracy of the search strings, which is open to debate. Nevertheless, the trends over time are systematic, suggesting that there is a shift in the explanatory language that is more than just an artifact of our search strings. Even though small, the share of papers with topological explanatory language starts to grow significantly after 2006. The share of papers with dynamic explanatory language is similarly low, but consistent and grows steadily since 1990. Papers with mechanistic explanatory language grew up to about 2002 and then seemed to stabilise. The graph also suggests that a large segment of papers (the remaining three quarters of the neuroscience literature) either does not use explanatory terms at all (e.g., it reports descriptive research), or uses explanatory terms not captured by our search strings.

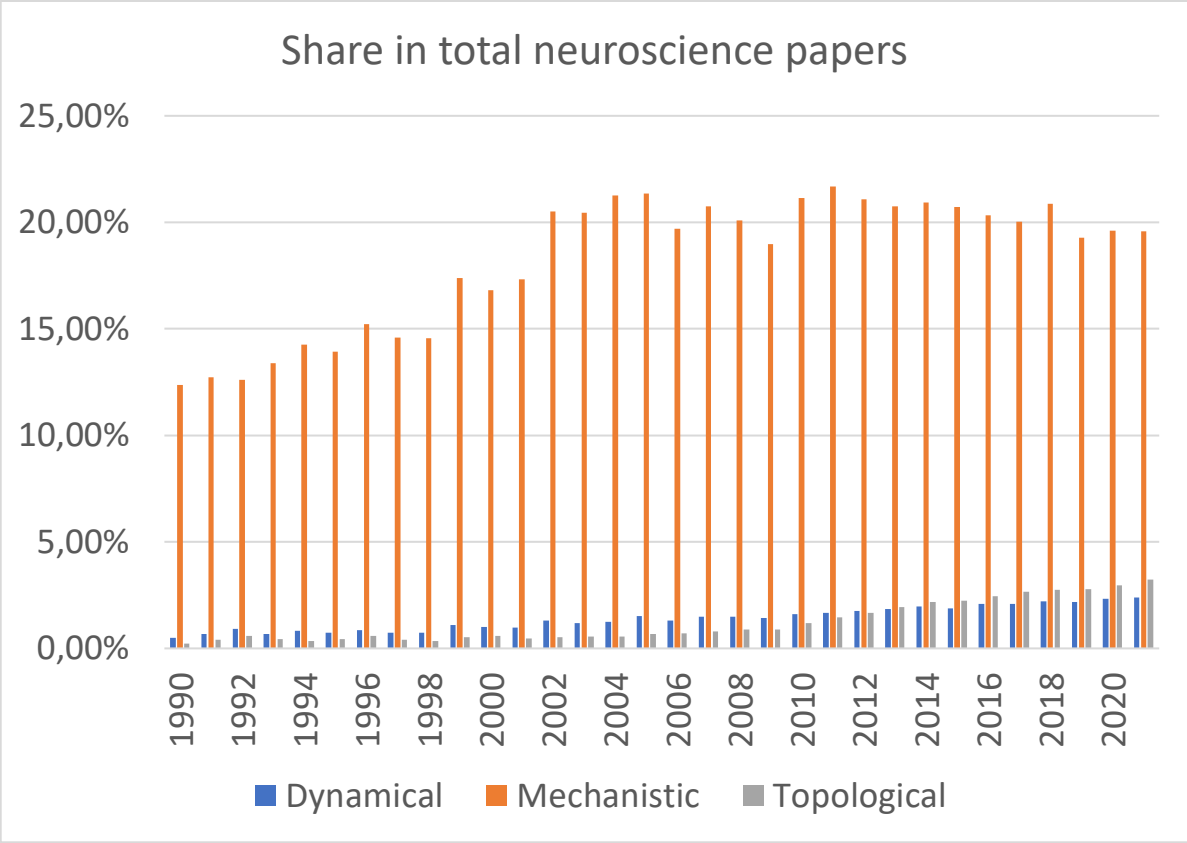

Figure 6. The ratio of mechanistic, topological and dynamical explanations in the total number of neuroscience papers from 1990-2020.

As an additional probe into the discriminatory power of our search strings, we analysed the overlap between the three sets. It has been suggested before (Petrovich and Viola 2022; Overton 2013) that the explanatory language used by scientists is not always entirely consistent and we may hence expect that some papers mix different explanatory terms. Figure 7 presents the number of papers in each set and the various overlaps between the sets. The largest overlap exists with papers with mechanistic explanatory language: about two-thirds of the papers in the dynamic and two-thirds of the papers in the topological set also contain mechanistic explanatory language. The overlap is smallest between papers in the topological and dynamic sets. However, an interesting trend can be observed if we represent the development of the summated overlaps between the three sets over time (Figure 8). Whereas the three sets nearly coincided up to about 2006, i.e., dominated by mechanistic explanatory language, after 2006 there is a steady trend towards less overlap: papers start to use more exclusively mechanistic, dynamic, or topological explanatory language.

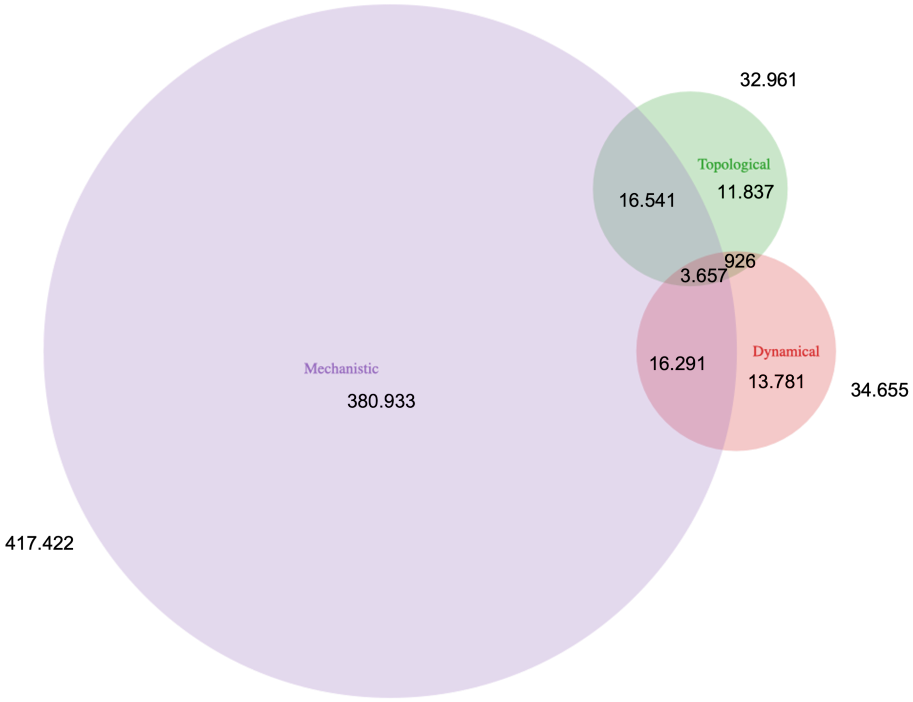

Figure 7. Overlap in search results (number of papers).

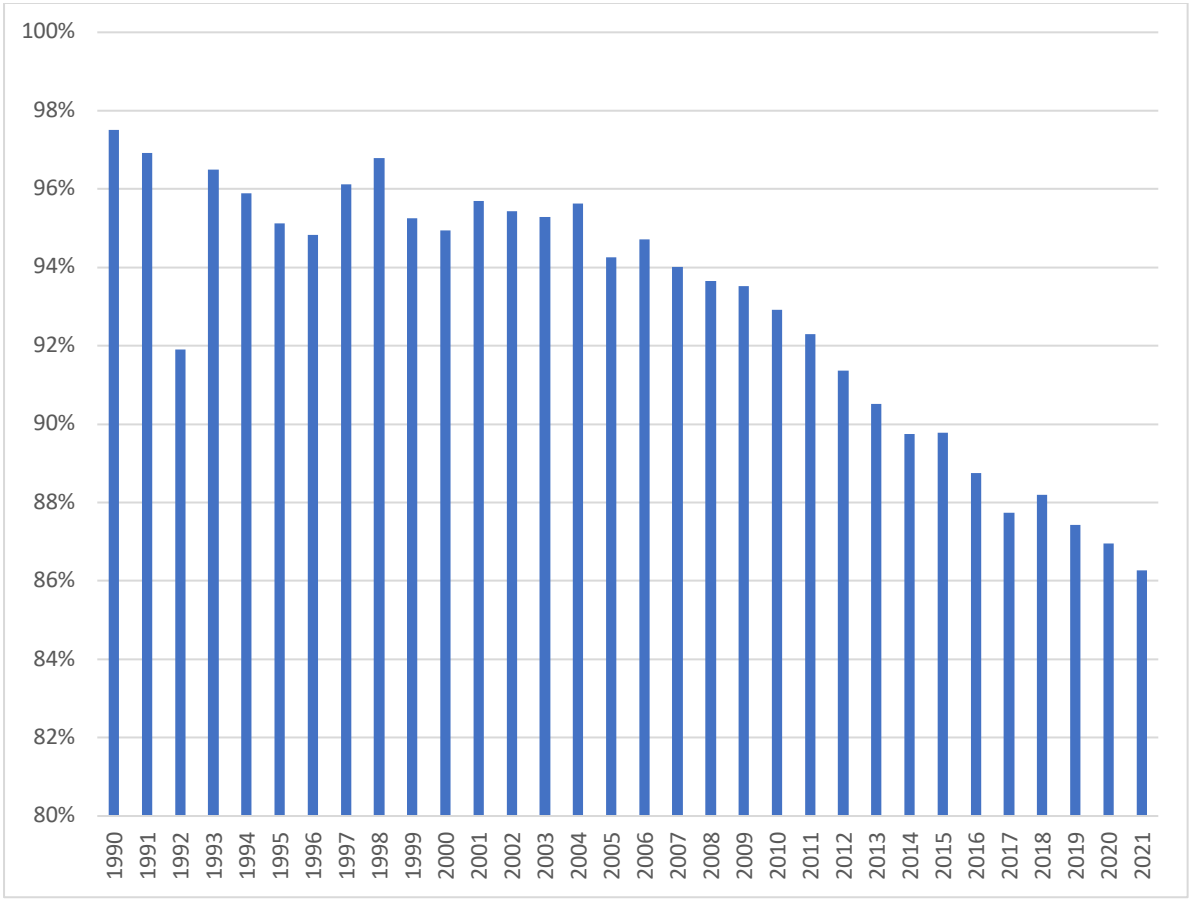

Figure 8: The total, summated overlap between dynamic, topological and mechanistic papers decreases over time.

5. Discussion: the explanatory landscape in neuroscience from 1990 to 2020, its trends, and the limits of bibliometric methodology

Our search strings returned only a limited set of the neuroscience literature, namely about a fifth. This may imply that we either missed a substantial part of the explanatory repertoire, or that a substantial part of the neuroscience literature does not use explanatory expressions (or both). Non-explanatory papers may be descriptive, i.e., they may be review articles, or papers that provide new data sets, describe new imaging techniques, or new tools for data analysis; or technical in nature, i.e., propose new experimental protocols, slight improvements on certain techniques, or in general be concerned with some form of "tinkering in the lab" (Bickle 2021). Of course, these non-explanatory uses would require further analysis. So, *pace* the new mechanists' claims that neuroscience is in the business of discovering mechanisms and *ipso facto* mechanistic explanations, it may be the case that a large part of neuroscience is in some other business than providing explanations, mechanistic or any other kind for that matter. Our study could not map out what that other business is, simply because the search strings were developed to identify specific and most typical explanatory linguistic patterns, and excluded less clear-cut expressions.

Having said that, within the fifth of the neuroscience literature that we analysed, our search strings suggest that mechanistic explanatory language is indeed predominant. Nevertheless, a significant number of dynamical and topological explanations papers exist, and their share slowly grows over time. The growth of topological explanation papers takes off around 2006. On the other hand, the number of papers that use the language of dynamical explanations show a steady growth without take-off points since the beginning of our corpus, the year 1990.

Unsurprisingly, the explanatory language is mixed. The topological and dynamical papers use mechanistic language too, which could be an artifact of noise generated by our search strings, or imprecise use of terms by neuroscientists, or a combination of multiple forms of explanation used in the same paper, probably a bit of all. However, the fact that the overlap of topological, dynamical and mechanistic language decreases over time, i.e., that they differentiate over time, may also indicate that loose mechanistic language was initially used as a placeholder for a more abstract non-mechanistic explanation, with which it is replaced over time as the ideas about dynamical and topological explanations start to develop and specialize.

The low number of topological explanation papers in our set (ca 0,5% of neuroscience papers) is to be expected, given that topological explanations were suggested relatively recently in the seminal paper by Sporns and colleagues (Sporns, Tononi, and Kötter 2005). To avoid contamination, we were also quite restrictive in the search terms we judged to be specifically indicative of topological explanations. On the other hand, topological explanations do use a more specific language in their *explanantia*, and because of that they are more discernible by our search strings. In contrast, mechanistic language is used more loosely (Dupré 2013; Woodward 2013; Ross 2021a; Kostić and Khalifa 2022), and so our (or any) search strings cannot discriminate between a genuine and platitudinous mechanistic explanatory language.

Our analysis has several limitations. One limitation is that we did not have access to a larger group of raters to do an extensive validation of our search strings in order to estimate false positives and false negatives. This more extensive validation, in our estimate, would

require a separate study, which is out of the scope of this paper. Our analysis therefore depends on our assessment of what counts as typical explanatory language, with the potential bias of one of the authors' previous work on one of the three types of explanation discussed in this paper. Another limitation is built into how searches work in the Dimensions database, with search strings producing a hit regardless of whether a string was used once or multiple times in the text. This includes casual as well as systematic use of explanatory language. A full-text analysis with more powerful tools, on a sample of neuroscience papers to keep it feasible, could provide more fine-grained results for a subset of papers. Such analysis could also use Natural Language Processing techniques, e.g., text data analyses that distinguish nouns from verbs. Perhaps, this method could better discern the explanatory language especially in papers that use mixed language.

Finding 'pure' dynamical, mechanistic and topological language might be possible, but it would require more precise assessment of every paper. Moreover, given that we are trying to detect different explanations by mapping explanatory language used by scientists, it would be possible to argue that these linguistic differences are a matter of conceptual sloppiness in the neuroscience literature that cannot be reduced to some overarching philosophical explanatory scheme. Nevertheless, these techniques can help to put theoretical debates in philosophy of science in an empirical perspective, in a systematic way, rather than based on hand-picked examples.

6. Conclusion

Explanatory language in neuroscience papers is not exclusively mechanistic. Our analysis has shown that a relatively small but growing share of neuroscience papers uses topological and dynamical terms to explain neural phenomena. We were also able to show that the explanatory repertoires in the neuroscience literature are differentiating over time: the explanatory language appears to become more exclusively either mechanistic, topological, or dynamical. Nevertheless, expressions of different types of explanatory language are regularly mixed in neuroscience papers.

Our study has shown that typical explanatory language can be identified by searching for particular word patterns. This approach could be expanded in several ways. First, similar word patterns could be identified for other types of explanation, such as the statistical language of correlation and association. Second, longer strings with more explanatory word patterns could be used to identify how explanation types are distributed throughout all of neuroscience. Our search strings returned only a fifth of neuroscience papers. Adding additional explanatory expressions will likely capture a larger share of the literature, although this would raise the share of false negatives and unwieldy data sets might then require smaller literature samples. Third, more refined natural language processing techniques could be used, such as techniques that analyse grammatical structures, e.g., distinguishing verbs from nouns.

We see no principled obstacle in applying our approach to life sciences in general, or, in fact, any other domain of science. The three types of explanation on which we focused in this paper are also used in other sciences, and repositories such as Dimensions.ai could provide corpora for them as well. Explanatory terms are most likely specific to each research field, but a similar two-step approach could be used, establishing dominant terms and then specific explanatory patterns for each field. However, in other fields, statistic or interpretative patterns may be more prevailing.

In studying the structure and dynamics of physics as a science, philosophers of science focus on its theories, i.e., how theories are formed, interrelated, and change over time. However, given the sheer diversity of explanatory styles in neuroscience, understanding its structure and dynamics seems rather a vexing task (Gold and Roskies 2008). In this paper we

focused on the neuroscience, and by mapping out different explanatory strategies in the large body of neuroscience literature, we have chosen an empirical approach to provide an overview of its structure and dynamics.

By looking at language use in a large corpus of literature, we provided empirical evidence about the explanatory landscape in neuroscience in general, and in that way, we avoid epistemic biases resulting from focusing solely on a limited and hand-picked examples that are typical of philosophical literature on scientific explanations. Our study demonstrates that the actual explanatory language in neuroscience is diversifying, rather than being exclusively or ever more mechanistic.

Appendix 1: lists of philosophical papers used to extract typical examples of mechanistic, topological and dynamical explanations

Philosophical accounts of mechanistic explanation

- Bechtel, William, and Robert C. Richardson. 2010. Discovering Complexity: Decomposition and Localization as Strategies in Scientific Research. MIT Press ed. Cambridge, Mass: MIT Press.
- Carl F. Craver. 2007. Explaining the Brain: Mechanisms and the Mosaic Unity of Neuroscience. Oxford : New York : Oxford University Press: Clarendon Press.
- Carl F., Craver and Lindley, Darden. 2013. In Search of Mechanisms: Discoveries across the Life Sciences. University of Chicago Press.
- Khalifa, Kareem, Farhan Islam, J. P. Gamboa, Daniel A. Wilkenfeld, and Daniel Kostić. 2022. "Integrating Philosophy of Understanding With the Cognitive Sciences." Frontiers in Systems Neuroscience 16 (March): 764708. https://doi.org/10.3389/fnsys.2022.764708.
- 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.
- Piccinini, Gualtiero, and Carl Craver. 2011. "Integrating Psychology and Neuroscience: Functional Analyses as Mechanism Sketches." Synthese 183 (3): 283–311. https://doi.org/10.1007/s11229-011-9898-4.

Philosophical accounts of topological explanation

- Khalifa, Kareem, Farhan Islam, J. P. Gamboa, Daniel A. Wilkenfeld, and Daniel Kostić. 2022. "Integrating Philosophy of Understanding With the Cognitive Sciences." Frontiers in Systems Neuroscience 16 (March): 764708. https://doi.org/10.3389/fnsys.2022.764708.
- Kostić, Daniel, Claus C Hilgetag, and Marc Tittgemeyer. 2020. "Unifying the Essential Concepts of Biological Networks : Biological Insights and Philosophical Foundations." Philosophical Transactions of the Royal Society B: Biological Sciences 375 (20190314): 1–5. https://doi.org/10.1098/rstb.2019.0314.
- Kostić, Daniel. 2018a. "Mechanistic and Topological Explanations: An Introduction." Synthese 195 (1): 1–10. https://doi.org/10.1007/s11229-016-1257-z.
- ———. 2018b. "The Topological Realization." Synthese 195 (1): 79–98. https://doi.org/10.1007/s11229-016-1248-0.
- Kostić, Daniel. 2019. "Unifying the Debates : Mathematical and Non-Causal Explanations." Perspectives on Science 2019, 27 (1): 1–6. https://doi.org/10.1162/posc.
- Kostić, Daniel. 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, November. https://doi.org/10.1007/s11229-021-03414-y.
- Ross, Lauren N. 2021. "Distinguishing Topological and Causal Explanation." Synthese 198 (10): 9803–20. https://doi.org/10.1007/s11229-020-02685-1.

Philosophical accounts of dynamical explanation

- Chemero, Anthony, and Michael Silberstein. 2008. “After the Philosophy of Mind: Replacing Scholasticism with Science*.” Philosophy of Science 75 (1): 1–27. https://doi.org/10.1086/587820.
- Favela, Luis H. 2020. “Dynamical Systems Theory in Cognitive Science and Neuroscience.” Philosophy Compass 15 (8). https://doi.org/10.1111/phc3.12695.
- ———. 2021. “The Dynamical Renaissance in Neuroscience.” Synthese 199 (1–2): 2103–27. https://doi.org/10.1007/s11229-020-02874-y.
- Gervais, Raoul. 2015. “Mechanistic and Non-Mechanistic Varieties of Dynamical Models in Cognitive Science: Explanatory Power, Understanding, and the ‘Mere Description’ Worry.” Synthese 192 (1): 43–66. https://doi.org/10.1007/s11229-014-0548-5.
- Khalifa, Kareem, Farhan Islam, J. P. Gamboa, Daniel A. Wilkenfeld, and Daniel Kostić. 2022. “Integrating Philosophy of Understanding With the Cognitive Sciences.” Frontiers in Systems Neuroscience 16 (March): 764708. https://doi.org/10.3389/fnsys.2022.764708.
- Stepp, Nigel, Anthony Chemero, and Michael T. Turvey. 2011. “Philosophy for the Rest of Cognitive Science.” Topics in Cognitive Science 3 (2): 425–37. https://doi.org/10.1111/j.1756-8765.2011.01143.x.
- Venturelli, A. Nicolás. 2016. “A Cautionary Contribution to the Philosophy of Explanation in the Cognitive Neurosciences.” Minds and Machines 26 (3): 259–85. https://doi.org/10.1007/s11023-016-9395-0.
- Verdejo, Víctor M. 2015. “The Systematicity Challenge to Anti-Representational Dynamicism.” Synthese 192 (3): 701–22. https://doi.org/10.1007/s11229-014-0597-9.

Appendix 2: lists of neuroscientific papers in three training corpora

Mechanistic training corpus

1. Aron, A., Fisher, H., Mashek, D. J., Strong, G., Li, H., & Brown, L. L. (2005). Reward, motivation, and emotion systems associated with early-stage intense romantic love. *Journal of neurophysiology*, *94*(1), 327-337.
2. Bahrami, B., Lavie, N., & Rees, G. (2007). Attentional load modulates responses of human primary visual cortex to invisible stimuli. *Current Biology*, *17*(6), 509-513.
3. Barsalou, L. W. (2008). Grounded cognition. *Annual review of psychology*, *59*(1), 617-645.
4. Dux, P. E., Ivanoff, J., Asplund, C. L., & Marois, R. (2006). Isolation of a central bottleneck of information processing with time-resolved fMRI. *Neuron*, *52*(6), 1109-1120.
5. Haynes, J. D., Sakai, K., Rees, G., Gilbert, S., Frith, C., & Passingham, R. E. (2007). Reading hidden intentions in the human brain. *Current biology*, *17*(4), 323-328.
6. Henson, R. (2006). Forward inference using functional neuroimaging: Dissociations versus associations. *Trends in cognitive sciences*, *10*(2), 64-69.
7. Kelso, S. (2010). Instabilities and phase transitions in human brain and behavior. *Frontiers in Human Neuroscience*, *4*, 23.
8. Knoch, D., Pascual-Leone, A., Meyer, K., Treyer, V., & Fehr, E. (2006). Diminishing reciprocal fairness by disrupting the right prefrontal cortex. *science*, *314*(5800), 829-832.
9. Logothetis, N. K. (2008). What we can do and what we cannot do with fMRI. *Nature*, *453*(7197), 869-878.
10. Logothetis, N. K., & Pfeuffer, J. (2004). On the nature of the BOLD fMRI contrast mechanism. *Magnetic resonance imaging*, *22*(10), 1517-1531.
11. Marder, E. (2015). Understanding brains: details, intuition, and big data. *PLoS biology*, *13*(5), e1002147.
12. Pecher, D., Zeelenberg, R., & Barsalou, L. W. (2004). Sensorimotor simulations underlie conceptual representations: Modality-specific effects of prior activation. *Psychonomic bulletin & review*, *11*(1), 164-167.
13. Poldrack, R. A. (2006). Can cognitive processes be inferred from neuroimaging data?. *Trends in cognitive sciences*, *10*(2), 59-63.
14. Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E., & Cohen, J. D. (2003). The neural basis of economic decision-making in the ultimatum game. *Science*, *300*(5626), 1755-1758.
15. Sengupta, P., & Samuel, A. D. (2009). Caenorhabditis elegans: a model system for systems neuroscience. *Current opinion in neurobiology*, *19*(6), 637-643.
16. Silvanto, J., Cowey, A., Lavie, N., & Walsh, V. (2005). Striate cortex (V1) activity gates awareness of motion. *Nature neuroscience*, *8*(2), 143-144.
17. Simmons, W. K., Ramjee, V., Beauchamp, M. S., McRae, K., Martin, A., & Barsalou, L. W. (2007). A common neural substrate for perceiving and knowing about color. *Neuropsychologia*, *45*(12), 2802-2810.
18. Sompolinsky, H. (2014). Computational neuroscience: beyond the local circuit. *Current opinion in neurobiology*, *25*, xiii-xviii.
19. Sumner, P., Tsai, P. C., Yu, K., & Nachev, P. (2006). Attentional modulation of sensorimotor processes in the absence of perceptual awareness. *Proceedings of the National Academy of Sciences*, *103*(27), 10520-10525.
20. Swartz, K. J. (2004). Towards a structural view of gating in potassium channels. *Nature Reviews Neuroscience*, *5*(12), 905-916.

Topological training corpus

1. Achard, S., Salvador, R., Whitcher, B., Suckling, J., & Bullmore, E. D. (2006). A resilient, low-frequency, small-world human brain functional network with highly connected association cortical hubs. *Journal of Neuroscience*, *26*(1), 63-72.
2. Adachi, Y., Osada, T., Sporns, O., Watanabe, T., Matsui, T., Miyamoto, K., & Miyashita, Y. (2012). Functional connectivity between anatomically unconnected areas is shaped by collective network-level effects in the macaque cortex. *Cerebral cortex*, *22*(7), 1586-1592.
3. Alexander-Bloch, A. F., Vértes, P. E., Stidd, R., Lalonde, F., Clasen, L., Rapoport, J., ... & Gogtay, N. (2013). The anatomical distance of functional connections predicts brain network topology in health and schizophrenia. *Cerebral cortex*, *23*(1), 127-138.
4. Betzel, R. F., Medaglia, J. D., Papadopoulos, L., Baum, G. L., Gur, R., Gur, R., ... & Bassett, D. S. (2017). The modular organization of human anatomical brain networks: Accounting for the cost of wiring. *Network Neuroscience*, *1*(1), 42-68.
5. Betzel, R. F., Gu, S., Medaglia, J. D., Pasqualetti, F., & Bassett, D. S. (2016). Optimally controlling the human connectome: the role of network topology. *Scientific reports*, *6*(1), 1-14.
6. Gu, S., Betzel, R. F., Mattar, M. G., Cieslak, M., Delio, P. R., Grafton, S. T., ... & Bassett, D. S. (2017). Optimal trajectories of brain state transitions. *Neuroimage*, *148*, 305-317.
7. Gu, S., Pasqualetti, F., Cieslak, M., Telesford, Q. K., Yu, A. B., Kahn, A. E., ... & Bassett, D. S. (2015). Controllability of structural brain networks. *Nature communications*, *6*(1), 1-10.
8. Helling, R. M., Petkov, G. H., & Kalitzin, S. N. (2019, January). 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* (pp. 1-6).
9. Hilgetag, C. C., & Goulas, A. (2016). Is the brain really a small-world network?. *Brain Structure and Function*, *221*(4), 2361-2366.
10. Honey, C. J., Thivierge, J. P., & Sporns, O. (2010). Can structure predict function in the human brain?. *Neuroimage*, *52*(3), 766-776.
11. Honey, C. J., Kötter, R., Breakspear, M., & Sporns, O. (2007). Network structure of cerebral cortex shapes functional connectivity on multiple time scales. *Proceedings of the National Academy of Sciences*, *104*(24), 10240-10245.
12. Hutchison, R. M., Womelsdorf, T., Allen, E. A., Bandettini, P. A., Calhoun, V. D., Corbetta, M., ... & Chang, C. (2013). Dynamic functional connectivity: promise, issues, and interpretations. *Neuroimage*, *80*, 360-378.
13. Kaiser, M., & Hilgetag, C. C. (2004). Edge vulnerability in neural and metabolic networks. *Biological cybernetics*, *90*(5), 311-317.
14. Kalitzin, S., Petkov, G., Suffczynski, P., Grigorovsky, V., Bardakjian, B. L., da Silva, F. L., & Carlen, P. L. (2019). Epilepsy as a manifestation of a multistate network of oscillatory systems. *Neurobiology of disease*, *130*, 104488.
15. Medaglia, J. D., Zurn, P., Armstrong, W. S., & Bassett, D. S. (2016). Mind control: frontiers in guiding the mind. http://www.arxiv.org>qbio>arXiv:1610.04134.
16. Meunier, D., Lambiotte, R., & Bullmore, E. T. (2010). Modular and hierarchically modular organization of brain networks. *Frontiers in neuroscience*, *4*, 200.
17. Mišić, B., Betzel, R. F., Griffa, A., De Reus, M. A., He, Y., Zuo, X. N., ... & Zatorre, R. J. (2018). Network-based asymmetry of the human auditory system. *Cerebral Cortex*, *28*(7), 2655-2664.
18. Ponten, S. C., Daffertshofer, A., Hillebrand, A., & Stam, C. J. (2010). The relationship between structural and functional connectivity: graph theoretical analysis of an EEG neural mass model. *Neuroimage*, *52*(3), 985-994.
19. Sporns, O., Tononi, G., & Kötter, R. (2005). The human connectome: a structural description of the human brain. *PLoS computational biology*, *1*(4), e42.
20. van den Heuvel, M. P., & Sporns, O. (2013). Network hubs in the human brain. *Trends in cognitive sciences*, *17*(12), 683-696.

Dynamical training corpus

1. Börgers, C., Epstein, S., & Kopell, N. J. (2008). Gamma oscillations mediate stimulus competition and attentional selection in a cortical network model. *Proceedings of the National Academy of Sciences*, *105*(46), 18023-18028.
2. Breakspear, M. (2017). Dynamic models of large-scale brain activity. *Nature neuroscience*, *20*(3), 340-352.
3. Bressler, S. L., & Kelso, J. S. (2001). Cortical coordination dynamics and cognition. *Trends in cognitive sciences*, *5*(1), 26-36.
4. Buhrmann, T., Di Paolo, E. A., & Barandiaran, X. (2013). A dynamical systems account of sensorimotor contingencies. *Frontiers in psychology*, *4*, 285.
5. Calcagnì, A., Lombardi, L., D'Alessandro, M., & Freuli, F. (2019). A state space approach to dynamic modeling of mouse-tracking data. *Frontiers in psychology*, *10*, 2716.
6. Connor, J. A. (1975). Neural repetitive firing: a comparative study of membrane properties of crustacean walking leg axons. *Journal of Neurophysiology*, *38*(4), 922-932.
7. Dale, R., & Bhat, H. S. (2018). Equations of mind: Data science for inferring nonlinear dynamics of socio-cognitive systems. *Cognitive Systems Research*, *52*, 275-290.
8. Dotan, D., & Dehaene, S. (2013). How do we convert a number into a finger trajectory?. *Cognition*, *129*(3), 512-529.
9. Freeman, J. B., & Ambady, N. (2011). When two become one: Temporally dynamic integration of the face and voice. *Journal of experimental social psychology*, *47*(1), 259-263.
10. Freeman, J. B., Dale, R., & Farmer, T. A. (2011). Hand in motion reveals mind in motion. *Frontiers in psychology*, *2*, 59.
11. Ijspeert, A. J., Nakanishi, J., Hoffmann, H., Pastor, P., & Schaal, S. (2013). Dynamical movement primitives: learning attractor models for motor behaviors. *Neural computation*, *25*(2), 328-373.
12. Jia, B., Gu, H. G., & Li, Y. Y. (2011). Coherence-resonance-induced neuronal firing near a saddle-node and homoclinic bifurcation corresponding to type-I excitability. *Chinese Physics Letters*, *28*(9), 090507.
13. Magnuson, J. S. (2005). Moving hand reveals dynamics of thought. *Proceedings of the National Academy of Sciences*, *102*(29), 9995-9996.
14. McKinstry, C., Dale, R., & Spivey, M. J. (2008). Action dynamics reveal parallel competition in decision making. *Psychological Science*, *19*(1), 22-24.
15. Maldonado, M., Dunbar, E., & Chemla, E. (2019). Mouse tracking as a window into decision making. *Behavior Research Methods*, *51*(3), 1085-1101.
16. Port, R. F. (2002). The dynamical systems hypothesis in cognitive science. *IULC Working Papers*, *2*(2).
17. Remington, E. D., Egger, S. W., Narain, D., Wang, J., & Jazayeri, M. (2018). A dynamical systems perspective on flexible motor timing. *Trends in cognitive sciences*, *22*(10), 938-952.
18. Shenoy, K. V., Sahani, M., & Churchland, M. M. (2013). Cortical control of arm movements: a dynamical systems perspective. *Annu Rev Neurosci*, *36*(1), 337-359.
19. Stephen, D. G., & Dixon, J. A. (2009). The self-organization of insight: Entropy and power laws in problem solving. *Journal of Problem Solving*, *2*(1), 72-102.
20. Tateno, T., Harsch, A., & Robinson, H. P. C. (2004). Threshold firing frequency–current relationships of neurons in rat somatosensory cortex: type 1 and type 2 dynamics. *Journal of neurophysiology*, *92*(4), 2283-2294.

Appendix 3: Mechanistic explanations: distinctive word patterns

explanatory relation	max words between	explanans	max words between	explanatory relation
Depends on	3	activity	3	controlling
Controlled by	1		3	control
Result of	1		1	crucial
Generate by	4		1	generate
underlying	3		1	affects
			1	dictates
			1	regulate
			4	produce
			5	modulate
			5	switch
			1	underlying
			4	Results in
			0	precedes
			2	elicit
			4	Responsible for

OR

explanatory relation	max words between	explanans	max words between	explanatory relation
Produced by	2	neuron	1	switch
Controlled by	0	neuronal	0	evoked
Performed by	0	neural		
Respond to	0			

Dynamical explanations: distinctive word patterns

explanatory relation	max words between	explanans	max words between	explanatory relation
depend on	2	dynamic	6	predict
driven by	1	dynamical	6	influence
governed by	2		1	constraint
dependent on	5		3	underlie
result from	5		6	determine
determined by	3		5	effect on
underlying	1		6	explain
generated by	2		4	produce
			2	lead to
			1	generate
			1	create
			1	work to

Topological explanations: distinctive word patterns

explanatory relation	max words between	explanans	max words between	explanatory relation
Contribution of	2	connectivity	4	confer*
Depends on	1	topolog*	9	constitut*
Determined by	4	architectur*	4	constrain*
Effect* of	2		1	determin*
Explanation of	3		2	enhance*
Influence of	2		1	explain*
Predicted	2		2	facilitate*
Role of	3		4	impact*
			5	Influence*
			6	play* role
			8	predict*
			3	relevant for
			5	responsible for
			6	shap*
			3	underlies

To avoid unnecessary repetition in the table, some of the terms are stemmed, e.g., explain* captures terms such as: explain, explains, explaining, or explained.

Conflict of Interest Statement

We declare that we have no conflict of interest.

References

Barabasi, Albert-Laszlo, and Reka Albert. 1999. "Emergence of Scaling in Random Networks" 286 (5439): 509–12. https://doi.org/10.1126/science.286.5439.509.

Barabási, Albert-László, and Zoltán N. Oltvai. 2004. "Network Biology: Understanding the Cell's Functional Organization." *Nature Reviews Genetics* 5 (2): 101–13. https://doi.org/10.1038/nrg1272.

Batterman, Robert W. 2010. "On the Explanatory Role of Mathematics in Empirical Science." *The British Journal for the Philosophy of Science* 61 (1): 1–25. https://doi.org/10.1093/bjps/axp018.

Batterman, Robert W., and Collin C. Rice. 2014. "Minimal Model Explanations." *Philosophy of Science* 81 (3): 349–76. https://doi.org/10.1086/676677.

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

Bickle, John. 2021. "Tinkering in the Lab." In *The Tools of Neuroscience Experiment*, by John Bickle, Carl F. Craver, and Ann-Sophie Barwich, 1st ed., 13–36. New York: Routledge. https://doi.org/10.4324/9781003251392-3.

Bonino, Guido, Paolo Maffezioli, Eugenio Petrovich, and Paolo Tripodi. 2022. "When Philosophy (of Science) Meets Formal Methods: A Citation Analysis of Early Approaches between Research Fields." *Synthese* 200 (2): 177. https://doi.org/10.1007/s11229-022-03484-6.

Carl F., Craver and Lindley, Darden. 2013. *In Search of Mechanisms: Discoveries across the Life Sciences*. University of Chicago Press.

Chemero, Anthony, and Michael Silberstein. 2008. "After the Philosophy of Mind: Replacing Scholasticism with Science." *Philosophy of Science* 75 (1): 1–27. https://doi.org/10.1086/587820.

Chirimuuta, M. 2014. "Minimal Models and Canonical Neural Computations: The Distinctness of Computational Explanation in Neuroscience." *Synthese* 191 (2): 127–53. https://doi.org/10.1007/s11229-013-0369-y.

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

Cupal, Jan, Stephan Kopp, and Peter F. Stadler. 2000. "RNA Shape Space Topology." *Artificial Life* 6 (1): 3–23. https://doi.org/10.1162/106454600568294.

Darrason, Marie. 2018. "Mechanistic and Topological Explanations in Medicine: The Case of Medical Genetics and Network Medicine." *Synthese* 195 (1): 147–73. https://doi.org/10.1007/s11229-015-0983-y.

Dupré, John. 2013. "I—John Dupré: Living Causes." *Aristotelian Society Supplementary Volume* 87 (1): 19–37. https://doi.org/10.1111/j.1467-8349.2013.00218.x.

Favela, Luis H. 2020. "Dynamical Systems Theory in Cognitive Science and Neuroscience." *Philosophy Compass* 15 (8). https://doi.org/10.1111/phc3.12695.

———. 2021. "The Dynamical Renaissance in Neuroscience." *Synthese* 199 (1–2): 2103–27. https://doi.org/10.1007/s11229-020-02874-y.

Fletcher, Samuel C., Joshua Knobe, Gregory Wheeler, and Brian Allan Woodcock. 2021. "Changing Use of Formal Methods in Philosophy: Late 2000s vs. Late 2010s." *Synthese* 199 (5–6): 14555–76. https://doi.org/10.1007/s11229-021-03433-9.
Gervais, Raoul. 2015. "Mechanistic and Non-Mechanistic Varieties of Dynamical Models in Cognitive Science: Explanatory Power, Understanding, and the 'Mere Description' Worry." *Synthese* 192 (1): 43–66. https://doi.org/10.1007/s11229-014-0548-5.
Glennan, Stuart. 2017. *The New Mechanical Philosophy*. First edition. Oxford: Oxford University Press.
Gold, Ian, and Adina L. Roskies. 2008. *Philosophy of Neuroscience*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195182057.003.0016.
Hitchcock, Christopher, and James Woodward. 2003. "Explanatory Generalizations Part II: Plumbing Explanatory Depth." *Noûs* 37 (2): 181–99. https://doi.org/10.1111/1468-0068.00435.
Huneman, Philippe. 2010. "Topological Explanations and Robustness in Biological Sciences," 213–45. https://doi.org/10.1007/s11229-010-9842-z.
Jones, Nicholaos. 2014. "Bowtie Structures, Pathway Diagrams, and Topological Explanation." *Erkenntnis* 79 (5): 1135–55. https://doi.org/10.1007/s10670-014-9598-9.
Khalifa, Kareem, Farhan Islam, J. P. Gamboa, Daniel A. Wilkenfeld, and Daniel Kostić. 2022. "Integrating Philosophy of Understanding With the Cognitive Sciences." *Frontiers in Systems Neuroscience* 16 (March): 764708. https://doi.org/10.3389/fnsys.2022.764708.
Kostić, Daniel, Claus C Hilgetag, Marc Tittgemeyer, Daniel Kostic, Daniel Kostic, Daniel Kostić, Original Scientific Paper, et al. 2019. "The Ultimate Articulation of the Account of Explanatory Understanding." *Topoi* 195 (1): 79–98. https://doi.org/10.1162/posc.
Kostić, Daniel. 2018a. "Mechanistic and Topological Explanations: An Introduction." *Synthese* 195 (1): 1–10. https://doi.org/10.1007/s11229-016-1257-z.
———. 2018b. "The Topological Realization." *Synthese* 195 (1): 79–98. https://doi.org/10.1007/s11229-016-1248-0.
———. 2020. "General Theory of Topological Explanations and Explanatory Asymmetry." *Philosophical Transactions of the Royal Society B: Biological Sciences* 375 (20190314): 1–8. http://dx.doi.org/10.1098/rstb.2019.0321.
———. (Forthcoming). "On the Role of Erotetic Constraints in Non-causal Explanations". *Philosophy of Science*.
Kostić, Daniel, Claus C Hilgetag, and Marc Tittgemeyer. 2020. "Unifying the Essential Concepts of Biological Networks : Biological Insights and Philosophical Foundations." *Philosophical Transactions of the Royal Society B: Biological Sciences* 375 (20190314): 1–5. https://doi.org/10.1098/rstb.2019.0314.
Kostić, Daniel, and Kareem Khalifa. 2022. "Decoupling Topological Explanations from Mechanisms." *Philosophy of Science*, April, 1–39. https://doi.org/10.1017/psa.2022.29.
Kostić, Daniel and Khalifa, Kareem. n.d. "The Directionality of Topological Explanations." *Synthese*, 23. https://doi.org/10.1007/s11229-021-03414-y.
Lange, Marc. 2013. "What Makes a Scientific Explanation Distinctively Mathematical?" *British Journal for the Philosophy of Science* 64 (3): 485–511. https://doi.org/10.1093/bjps/axs012.
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.

Malaterre, Christophe, Jean-François Chartier, and Davide Pulizzotto. 2019. “What Is This Thing Called *Philosophy of Science* ? A Computational Topic-Modeling Perspective, 1934–2015.” *HOPOS: The Journal of the International Society for the History of Philosophy of Science* 9 (2): 215–49. https://doi.org/10.1086/704372.

Mizrahi, Moti, and Michael Adam Dickinson. 2022a. “Philosophical Reasoning about Science: A Quantitative, Digital Study.” *Synthese* 200 (2): 138. https://doi.org/10.1007/s11229-022-03670-6.

———. 2022b. “Is Philosophy Exceptional? A Corpus-Based, Quantitative Study.” *Social Epistemology*, August, 1–18. https://doi.org/10.1080/02691728.2022.2109529.

No Title. n.d.

Overton, James A. 2013. “‘Explain’ in Scientific Discourse.” *Synthese* 190 (8): 1383–1405. https://doi.org/10.1007/s11229-012-0109-8.

Petrovich, Eugenio, and Marco Viola. 2022. “The ‘Cognitive Neuroscience Revolution’ Is Not a (Kuhnian) Revolution. Evidence from Scientometrics.” *Rivista Internazionale Di Filosofia e Psicologia* 13 (2): 142–56. https://doi.org/10.4453/rifp.2022.0013.

Piccinini, Gualtiero, and Carl Craver. 2011. “Integrating Psychology and Neuroscience: Functional Analyses as Mechanism Sketches.” *Synthese* 183 (3): 283–311. https://doi.org/10.1007/s11229-011-9898-4.

Rathkopf, Charles. 2015. “Network Representation and Complex Systems.” *Synthese*, 1–26. https://doi.org/10.1007/s11229-015-0726-0.

Rice, Collin. 2021. *Leveraging Distortions: Explanation, Idealization, and Universality in Science.* MIT Press.

Ross, Lauren N. 2021a. “Causal Concepts in Biology: How Pathways Differ from Mechanisms and Why It Matters.” *The British Journal for the Philosophy of Science* 72 (1): 131–58. https://doi.org/10.1093/bjps/axy078.

———. 2021b. “Distinguishing Topological and Causal Explanation.” *Synthese* 198 (10): 9803–20. https://doi.org/10.1007/s11229-020-02685-1.

Sporns, Olaf, Giulio Tononi, and Rolf Kötter. 2005. “The Human Connectome: A Structural Description of the Human Brain.” *PLoS Computational Biology* 1 (4): 0245–51. https://doi.org/10.1371/journal.pcbi.0010042.

Stadler, Bärbel M R, and Peter F Stadler. 2004. “The Topology of Evolutionary Biology.” *Modelling in Molecular Biology*, 267–86. https://doi.org/10.1007/978-3-642-18734-6_12.

Stepp, Nigel, Anthony Chemero, and Michael T. Turvey. 2011. “Philosophy for the Rest of Cognitive Science.” *Topics in Cognitive Science* 3 (2): 425–37. https://doi.org/10.1111/j.1756-8765.2011.01143.x.

Venturelli, A. Nicolás. 2016. “A Cautionary Contribution to the Philosophy of Explanation in the Cognitive Neurosciences.” *Minds and Machines* 26 (3): 259–85. https://doi.org/10.1007/s11023-016-9395-0.

Verdejo, Víctor M. 2015. “The Systematicity Challenge to Anti-Representational Dynamicism.” *Synthese* 192 (3): 701–22. https://doi.org/10.1007/s11229-014-0597-9.

Vernazzani, Alfredo. 2019. “The Structure of Sensorimotor Explanation.” *Synthese* 196 (11): 4527–53. https://doi.org/10.1007/s11229-017-1664-9.

Walsh, D. M. 2014. “Variance, Invariance and Statistical Explanation.” *Erkenntnis*, 469–89. https://doi.org/10.1007/s10670-014-9680-3.

Walsh, D M, T Lewens, and A Ariew. 2002. “The Trials of Life: Natural Selection and Random Drift.” *Philosophy of Science* 69 (3): 452–73. https://doi.org/10.1086/342454.

Watts, Duncan, and Steven H Strogatz. 1998. “Collective Dynamics of ‘small-World’ Networks.” *Nature* 393 (6684): 440–42. https://doi.org/10.1038/30918.

Weiskopf, Daniel A. 2011. “Models and Mechanisms in Psychological Explanation.” *Synthese* 183 (3): 313–38. https://doi.org/10.1007/s11229-011-9958-9.
Woodward, James. 2013. “II—James Woodward: Mechanistic Explanation: Its Scope and Limits.” *Aristotelian Society Supplementary Volume* 87 (1): 39–65. https://doi.org/10.1111/j.1467-8349.2013.00219.x.
Woodward, James, and Christopher R Hitchcock. 2003. “Explanatory Generalizations, Part I: A Counterfactual Account.” *Nous* 37 (1): 1–24. https://doi.org/10.1111/1468-0068.00426.