Chapter 5: Evolving Concepts of "Hierarchy" in Systems Neuroscience

Daniel Burnston[1] and Philipp Haueis[2]*

* Authors appear in alphabetical order and contributed equally to this article.

ABSTRACT: The notion of "hierarchy" is one of the most commonly posited organizational principles in systems neuroscience. To this date, however, it has received little philosophical analysis. This is unfortunate, because the general concept of hierarchy ranges over two approaches with distinct empirical commitments, and whose conceptual relations remain unclear. We call the first approach the "representational hierarchy" view, which posits that an anatomical hierarchy of feed-forward, feed-back, and lateral connections underlies a signal processing hierarchy of input-output relations. Because the representational hierarchy view holds that unimodal sensory representations are subsequently elaborated into more categorical and rule-based ones, it is committed to an increasing degree of abstraction along the hierarchy. The second view, which we call "topological hierarchy," is not committed to different representational functions or degrees of abstraction at different levels. Topological approaches instead posit that the hierarchical level of a part of the brain depends on how central it is to the pattern of connections in the system. Based on the current evidence, we argue that three conceptual relations between the two approaches are possible: topological hierarchies could *substantiate* the traditional representational hierarchy, *conflict* with it, or contribute to a *plurality* of approaches needed to understand the organization of the brain. By articulating each of these possibilities, our analysis attempts to open a conceptual space in which further neuroscientific and philosophical reasoning about neural hierarchy can proceed.

KEYWORDS: hierarchy, systems neuroscience, representation, topology, abstraction

[1] Department of Philosophy, Tulane Brain Institute, Tulane University, New Orleans, USA

e-mail: dburnsto@tulane.edu

[2] Department of Philosophy, Bielefeld University, Bielefeld, Germany

e-mail: philipp.haueis@uni-bielefeld.de

1. Introduction

Scientific concepts evolve over time. As researchers generate new data and explore an increasing number of related yet subtly different phenomena, concepts frequently acquire novel connotations and expand their reference to novel properties. What is often left is a patchwork of multiple meanings and uses operating under the guise of a univocal concept. Because they result from the exploration of related phenomena, patchwork concepts are polysemous, i.e. they have multiple related meanings (as opposed to ambiguous words, whose distinct meanings are unrelated, cf. Sennet 2016). Recent case studies in the physical and life sciences suggest that such polysemous patchwork concepts help researchers to describe distinct but related phenomena efficiently (Wilson, 2006), classify properties at different scales (Bursten, 2016), or integrate seemingly incompatible uses of a concept in theoretically fruitful ways (Novick, 2018; Haueis, 2018). Scholars within this literature have primarily focused on patchworks as a descriptive claim about concept development within science, and on the positive contributions of patchwork concepts to the projects researchers pursue.

We agree on the descriptive claim that polysemous patchwork concepts are a pervasive feature of scientific language. We suggest, however, that the normative status of concepts with multiple related meanings is a genuinely open issue. Why *should* patchwork concepts be developed during investigation? We suggest that although patchwork concepts allow the investigation of phenomena that are closely related, they do not determine the exact relationship between them. Thus, how any two meanings of a conceptual patchwork are properly related depends on the exact relationship between the phenomena they describe. The meanings may overlap if the phenomena they describe are identical. Or the meanings may diverge if the phenomena they describe are distinct. Or one meaning may be an accurate description of some phenomenon, while another is not.

So, developing a patchwork concept allows for investigation of closely related phenomena to proceed without proscribing the relationship between them. But there is a downside to this process – concepts often change "silently," with new connotations emerging in the course of investigation, and without those differences explicitly acknowledged. The appropriate normative attitude to patchworks

involves a commitment to explicitly cashing out the distinct aspects of the patchwork, so that the relationships between the phenomena they describe can be investigated empirically.

We explore these issues by analyzing the concept of "hierarchy" in systems neuroscience. As we will outline, the idea that the brain is hierarchically organized has had a long and influential history in the field. Neuroscientists have just begun to recognize, however, that the concept comprises multiple distinct connotations that are often not distinguished (Hilgetag and Goulas 2020). We analyze (i) why the patchwork has developed, (ii) the different connotations it currently comprises, and (iii) the different possible relationships between connotations within the patchwork. Our analysis thus advances both descriptive and normative aspects of the patchwork approach, and provides clarity on a conceptually difficult issue within the neurosciences.

Posits of hierarchical organization are practically ubiquitous in systems neuroscience, but we contend that the concept currently ranges over two broadly distinct approaches with different core commitments. The first, which we call the "representational hierarchy" view, is extremely influential in the field. The representational hierarchy view posits an anatomical hierarchy of feed-forward, feedback, and lateral connections which underlies a sequence of input-output relations between brain areas. During this process, simple, unimodal sensory representations are subsequently elaborated into categorical, multimodal, and rule-based ones. The second, much newer view we call the "topological" approach, which is primarily based on the notion of *centrality*. A brain area is at a higher hierarchical level if it has more widespread influence on the network of brain areas. The topological approach primarily employs tools from graph theory and also focuses on an area's temporal contribution to evolving brain dynamics.

Although the two views are deeply intertwined in current systems neuroscience, we suggest that they have distinct central commitments.[3] The representational hierarchy view is committed to specific

---

[3] Hilgetag and Goulas (2020) distinguish four instead of two senses of hierarchy. Although a detailed comparison of both taxonomies is beyond the scope of this chapter, we think that that their definitions of hierarchy as laminar projection patterns and as spatial gradients of structural features shares the commitments of what we call the "representational approach" to hierarchy (section. 4.2, Fig. 3).

hypotheses about the representational roles of brain parts at distinct hierarchical levels. The topological view has no such commitments. Establishing the distinction between the views allows us to ask about the relationship between them. We consider three possibilities. First, the *substantiation* view suggests that the topological hierarchies provide a more detailed view of the anatomical underpinnings of representational hierarchies. Second, the *conflict* view states that the topological approach is a potential *replacement* for the representational view. Finally, there are several possible varieties of *pluralism*, which hold that the representational and topological approaches are mutually compatible depictions of distinct aspects of brain organization.

Our discussions will be internal to the neuroscience literature, but we hasten to add that frameworks in cognitive science and philosophy of mind often employ the representational hierarchy view. Consider debates about cognitive penetration and higher-level content, which implicitly presume that "lower-level" perception involves representation of simpler perceptual features. The question is whether perception can represent more abstract categories at a "higher" level of processing (Orlandi, 2010), and whether this is due to "top-down" influence from brain parts that represent concepts (Vetter & Newen, 2014). Or consider predictive coding models, which often cite hierarchical representations in the brain to argue that feed-back connections deliver predictions based on higher-level generalizations to sensory areas (Bastos et al. 2012; Hohwy, 2013). Each of these positions is broadly committed to the representational hierarchy view, and thus entails either the substantiation view or some variety of pluralism. Given that the conflict view is also possible, this cannot simply be assumed.

We proceed as follows. In section 2, we introduce the representational hierarchy approach, and in section 3 the topological approach. Section 4 then articulates the substantiation, conflict, and pluralist views. In section 5, we consider studies of the rich club phenomenon within the topological approach as a test case for the different views of the relationship. Section 6 concludes.

Similarly, we think that their definitions of hierarchy in terms of topological projection sequences and as multilevel modular networks share the commitments of what we call the "topological approach" (section 3., Fig. 4).

2. The representational approach to hierarchy

The traditional – and, by far, the common – approach takes anatomical connections in the cortex to reveal a hierarchical organization in patterns of feedforward and feedback connections. The locus classicus of this approach is Felleman and Van Essen (1991). Drawing on histological data, they posited definitions of hierarchical "level" as depicted in Figure 1.
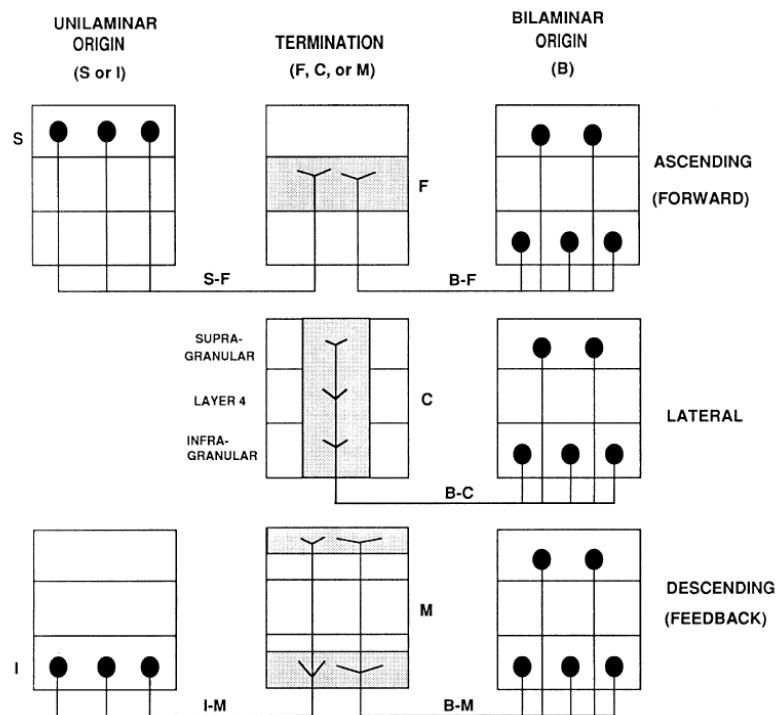

Fig. 1. Definitions of hierarchical relationships. From Felleman and Van Essen (1991).

The first row of Fig 1 shows two connection patterns that, according to Felleman and Van Essen's framework, count as *ascending* or feed-forward: either the connection begins in "supragranular" layers (layers 1–3 of cortex, left panel) and terminate in layer 4 (middle panel). Or it originates in both supra- and "infragranular" layers (5-6, right panel) and terminates in layer 4 (middle panel). The second row of Fig. 1 shows that lateral connections begin in both supra- and

infragranular layers and terminate in all layers. The third row shows that descending/feedback connections also begin at supra- and infra-granular layers but terminate in all but layer 4. This scheme can be used to classify different parts of the brain into hierarchical levels, based purely on anatomical connectivity. A given area A is at a higher hierarchical level than another area B if A received only feed-forward connections from B, and B received only feedback connections from A. Two areas are on the same hierarchical level if (i) they share only lateral connections, or (ii) they have similar patterns of feed-forward and feedback connections to already established levels. Based on this scheme, Felleman and Van Essen constructed a hierarchical description of the visual cortex comprising of 10 levels. The overall picture, as shown in Figure 2, has been extraordinarily influential, and is often taken as an exemplar for describing organization in the brain (Bechtel, 2008, ch. 3).
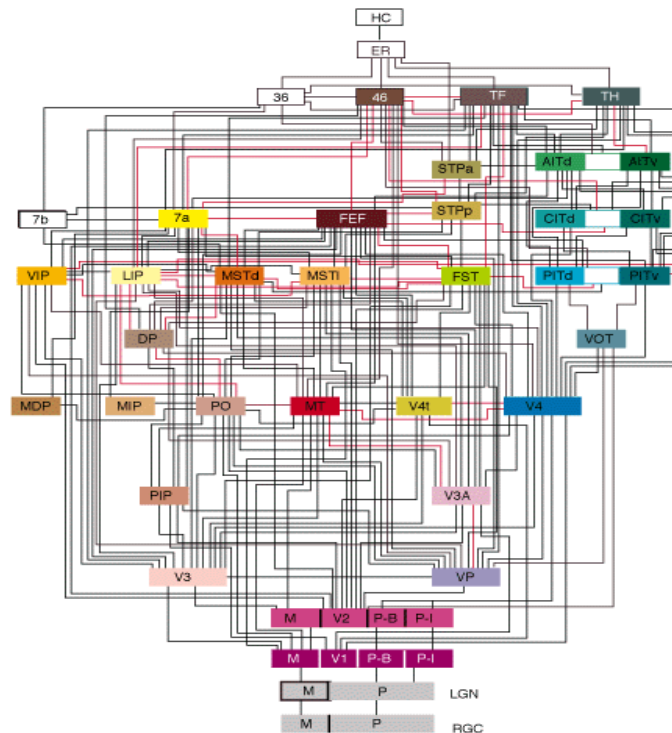

Fig. 2. The hierarchical wiring diagram of the macaque visual cortex. From Felleman and Van Essen (1991).

While their analysis was based on anatomy, Felleman and Van Essen did not shy away from applying a functional and representational interpretation of their framework:

> The physiological properties of any given cortical neuron will, in general, reflect many descending as well as ascending influences. Nevertheless, the cell may represent a well-defined hierarchical position in terms of the types of information it represents explicitly and the way in which that information is used." (Felleman & Van Essen, 1991, p. 32).

On this view, the hierarchical position of a brain area connotes a functional and representational specificity: occupying a specific place in the hierarchy involves representing certain types of information and representing that information for further use elsewhere in the system. This approach is generally seen as a way of extending Hubel and Wiesel (1962), who showed how patterns of anatomical connectivity can combine to produce new functional representations. In Fig. 3, three "simple cells" (upper right) represent the orientation of an edge at a particular place in the visual field (small triangles and crosses on the left). The simple cells then forward these representations to a single "complex" cell (lower right). The complex cell will then represent the orientation *wherever* it occurs across the receptive fields of the simple cells (dotted rectangle, right).
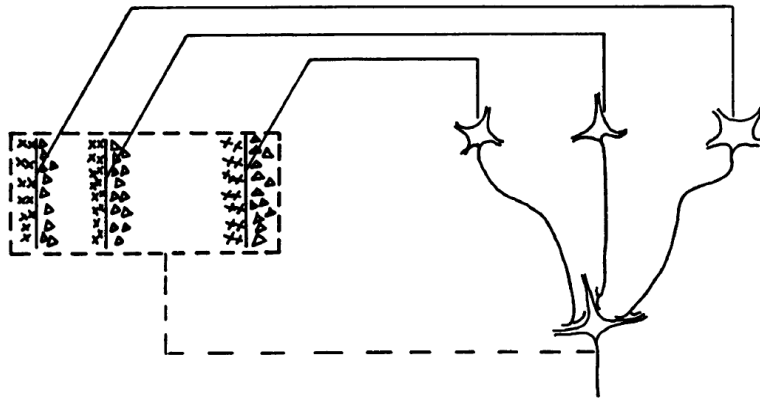

Fig. 3. The hierarchical logic explaining complex receptive field properties of V1 neurons in cat cortex. From Hubel and Wiesel (1962).

Figure 3 points to principles of processing within the hierarchy – *specific* information is passed along feedforward pathways, and then is represented *more abstractly* by higher levels in the hierarchy. The representational hierarchy view extends this logic to the rest of the visual system: lower levels of the hierarchy (including V1, V2, and V3) represent extremely simple features (such as orientation, wavelength, and displacement) at specific places in the visual field. At higher levels of the hierarchy more abstract information is represented. Within the dorsal stream for instance, MT represents general patterns of motion whereas V1 represents only local displacement. Within the ventral stream, a dedicated part of V4 represents categories of color whereas V1 represents only wavelength. A different part of V4 represents complex shapes rather than V1's representation of local orientation. Higher-level areas such as the inferotemporal cortex represent objects when they belong to a *category*, such as faces or hands, despite variation in their specific lower-level feature values (Gross, Rocha-Miranda, & Bender, 1972). Due to its view of functional and representational organization, Burnston (2016a, 2016b) has dubbed this view the "modular functional hierarchy" (MFH) picture of visual cortex organization.

Early on, it was noted that there were serious empirical shortcomings with Felleman and Van Essen's approach. In particular, many different possible attributions of hierarchical levels were compatible with the known data (Hilgetag, O'Neill and Young, 1996). Still, the MFH view in general has had an astounding effect on the field of systems neuroscience and has extended well beyond the visual system. Here is a small set of examples.

First, the MFH view has intersected with computer vision to produce a picture of how categorical perception comes about. Influential approaches by Poggio (e.g., Riesenhuber & Poggio, 1999) and Ullman (2007) have implemented feedforward networks that begin with representations of simple features and subsequently represent more abstract categories. Ullman's hierarchy is based explicitly on representing fragments of lesser complexity at lower levels, and then, on the basis of these, representing the category of the object at a subsequent stage of processing. These feedforward approaches, however, are also getting increasingly replaced by recurrent deep

neural network architectures in computational approaches to visual object recognition.

Second, the MFH view has been used to analyze *other* sensory systems. The idea is that analogues to the simple features of the visual system can be found, and that these will be represented at lower levels of an anatomical hierarchy that works similarly to the one in the visual system. Such views have been proposed for both the olfactory and the auditory system (Savic, Gulyas, Larsson, & Roland, 2000; Wessinger et al., 2001).

Third, the MFH view is taken to describe motor systems. Interestingly, however, in these systems the primary direction of influence is taken to be the reverse of sensory systems. Abstract goal representations are encoded at the top of the hierarchy, localized to areas such as the premotor cortex and the inferior parietal lobule (Grafton and Hamilton, 2007; for further discussion see Uithol, Burnston, & Haselager, 2014), and these are subsequently expanded into a representation of the detailed object properties and motor kinematics needed to attain the outcome. Grafton and Hamilton (2007) explicitly analogize this to the kind of sequential representational hierarchy in the visual system (cf. Haggard, 2005).

Finally, a hierarchy of abstraction for action control is often posited to explain the organization of the dorsolateral prefrontal cortex. In a classic fMRI study, Koechlin et al. (2003) had subjects perform a series of successively more complex actions. In the simplest case, subjects had to perform a motor action in response to a visual cue. In the harder case, the stimulus-response associations shifted, depending on a second cue. In the hardest case, the overall pattern of associations between cues and sensorimotor associations changed depending on still another cue. The structure of this task is hierarchical, with sensorimotor associations nested under conditions, and conditions nested under episodes. More anterior areas of the dlPFC were activated with increasing hierarchical nesting of the needed cognitive control. Badre and D'Esposito (2010) take these and similar results to show that anterior areas are involved in the employment of abstract rules.

The representational hierarchy approach thus supports an overall view of brain function. On this picture, unimodal and motor cortices each embody a representational hierarchy. The outputs of perceptual systems are brought together in "association" cortices, including frontal and parietal areas (Mesulam, 1998). Multimodal information is processed according to rules in executive control areas such as the dlPFC, and motor systems implement goals via specific representations of motor kinematics. Thus, the representational hierarchy view posits principles based on increasing abstraction for both unimodal and association cortices and for the overall functional architecture of the brain. In the next section, we discuss topological hierarchies, before moving on to discuss potential relationships between the two views.

3. Topological approaches to hierarchy

Topological approaches to hierarchy use the mathematical tools of graph theory to describe the brain as a network consisting of nodes (e.g., brain areas or individual neurons) and edges (e.g., axonal connections, fiber pathways). Topological approaches are distinct from the representational hierarchy view because topological hierarchies quantitatively describe the potential influence of a given node on the system, rather than positing a specific type or degree of abstraction of the information it processes. This focus on potential influence makes topological approaches neutral to the representational functions of given parts of the brain. After introducing the graph theoretical concepts used to describe brains as hierarchical, we describe two ways to specify topological hierarchy and argue that neither of them is committed to representational functions. This neutrality prepares our argument that different relations between representational and topological approaches are possible (section 4).

The representational neutrality we emphasize here intersects with, but is distinct from, several recent discussions in the philosophy of science literature, which attempt to address the relationships between network-based explanations and mechanistic explanations. Several authors have stressed the *distinctness* of these forms of explanation and debated the relationship between them. In particular,

those who think that topological explanations are entirely distinct from mechanistic ones tend to stress their abstraction (Huneman 2010), or the fact that they describe *global* properties of systems, rather than local causal interactions (Kostić 2016, Rathkopf, 2018).

We also rely on the abstractness of graph-theoretic explanation in articulating the difference between distinct conceptions of hierarchy. Graph-theoretical descriptions are neutral with respect to the representational role of different hierarchical levels because they are not committed to a particular way of functionally typing the causal interactions in the brain. However, we do not take this itself to show that topological explanation is always global, or is in conflict with mechanistic explanation in general. This is compatible with the *explanation* of particular phenomena invoking local causal interactions as well as global organizational properties. We take no particular stand on the issue here (but see Burnston 2019).

The graph-theoretical notion of hierarchy is based on the concept of *centrality* (see van den Heuvel and Sporns 2013 for review and further references). A node is at a higher hierarchical level if it is more central to the overall connectivity of the network, and at a lower level if it is more peripheral. Centrality can be analyzed in different ways. One notion is simply *degree* – a node with a large number of connections (measured as percentage of actual out of possible connections) will have a large influence in the network. An example of a degree measurement is given in the left panel of figure 4 below. A second notion of centrality is *betweenness centrality*, i.e. how many shortest paths between any two nodes pass through the node of interest. Nodes with high betweenness centrality are crucial for mediating interactions across the entire network. Finally, the *clustering coefficient* of a node measures the degree to which the node's connections are themselves connected. It is measured as the proportion of actual out of possible edges between nodes that are connected to the node of interest.
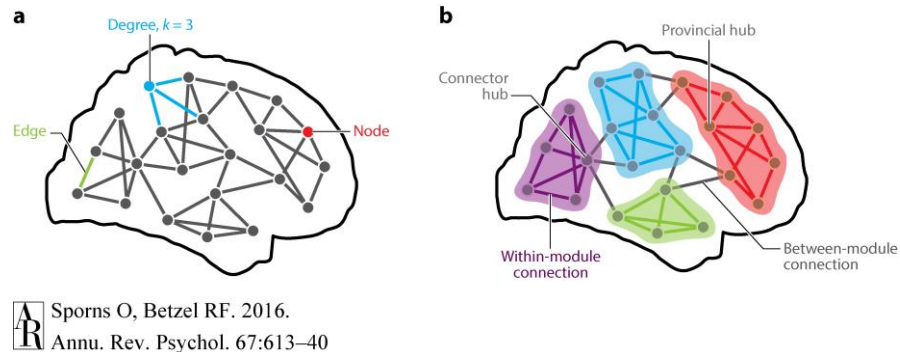
Fig. 4. Hierarchical measurements in the topological approach (from Sporns & Betzel, 2016).

Centrality measures can be used to describe the overall properties of the network as well as particular nodes, particularly in how network organization is distributed amongst *modules* and *hubs*. Nodes in a module are more connected amongst each other than to nodes outside the module (Sporns and Betzel 2016). Consequently, these within-module nodes will influence each other more directly than other nodes. Hubs are nodes (or groups of nodes) which score high on one or multiple centrality measures, which are usually correlated (van den Heuvel and Sporns 2013). A hub with a high clustering coefficient is likely to connect several modules, and thus provide information transfer across otherwise segregated subsystems ("connector hubs"; Fig. 4 above). A node can also serve as a hub primarily within, rather than between modules, by mostly connecting to other nodes in the same module ("provincial hubs"; Fig 4 above). The extent to which networks exhibit modularity and contain hubs gives a helpful characterization of their overall capacity to process information. When a network contains primarily modules with a smaller number of hubs, it can maximize both localized information processing through within-module connections, and information integration across the network through hub-mediated connections (Sporns, 2011).

From these definitions one can already see why "influence on the network" is the primary notion for any topological approach to neu-

ral hierarchies.[4] If nodes are defined as brain areas, then activity in a highly central area will influence activity in many other areas, and thus shape the global behavior of the network. A topological hierarchy description of the brain is generated by applying the aforementioned centrality measures to anatomical or functional connectivity data. Some of these datasets include the kind of histological data cited in the discussion of Felleman and Van Essen (e.g., the CoCoMac database), but have been updated to include more complete data about neural connections. Functional connectivity is, basically, a measure of the statistical correlation in activity between brain areas over time (it can be measured in different ways, and we won't go into to go into the details here; see Haueis 2012 for discussion). Here we give some specific examples where researchers have employed the topological approach to hierarchy to make sense of brain organization.

An early example of the topological approach, as applied to anatomical connectivity, is from da Costa and Sporns (2005), who used degree and clustering coefficient to study the hierarchical organization of the macaque visual system. Their analysis was based on how closely a starting brain area (a "reference node") was connected to the rest of the system. They thus analyzed each area in terms of *degree* distance. From a given reference node, for instance, they asked how many other nodes it connected to with only one synaptic connection, how many at two synapses distant, etc. They defined "levels" as degree measures at distinct synaptic distances, and showed that six areas in the visual system, predominantly in the dorsal stream, connect to more than half of the rest of the visual system at the first hierarchical level. These areas thus have the most direct

---

[4] While we focus on the influence notion of hierarchy, other network investigations employ a more compositional notion of hierarchy as well. For instance, researchers also talk of "hierarchy" if network structure is self-similar, e.g. when smaller modules are nested within larger modules (Hilgetag and Goulas 2020). While it may be interesting to analyze how such "encapsulation hierarchies" relate to compositional hierarchies in the mechanistic literature (Craver 2007, ch. 5), in the following we assume that systems neuroscientists studying encapsulation hierarchies are usually interested in its implications for neural signaling, i.e. on how influential a brain part is within the network (Müller-Linow et al. 2008, Sporns and Betzel 2016).

influence on many other areas of the visual network. Ventral stream areas predominantly connect to other nodes at the second and third hierarchical level, which means that their influence is less central. An exception was area V4, which is in the ventral stream, but had similarly high degree measures at a degree distance of one. (We will discuss their analysis of clustering coefficients in section 4.)

Centrality-based analyses of structural connectivity have also been used to study the entire brain. For instance, Zamora-Lopéz et al. (2010) used degree and betweenness centrality to determine the distribution of hubs in the cat cortex. Their analysis revealed that most nodes with high betweenness centrality lie in frontal and limbic cortex, and only few in sensory cortices. In addition to purely structural connectivity in cats and primates, Centrality measures have been applied to functional connectivity in humans. Meunier et al. (2009) used degree and modularity measures to describe functional connectivity data recorded with fMRI during the experimental resting state. They showed that only five percent of the nodes qualify as hubs that connect several modules, suggesting that these areas of the brain are particularly central. In particular, they showed that the areas of the "default-mode" network (DMN), which have been shown to be highly active during rest, are themselves both highly interconnected (thus forming a module) *and* highly connected to the rest of the brain (thus forming a hub). We discuss the DMN more thoroughly in subsequent sections. [5]

Both anatomical and functional connectivity measures are importantly *static* – they describe the state of the brain as a constant within a period of time (e.g., during rest). But network measures can also be used to describe dynamics. In the temporal domain, topological hierarchies posit that nodes with activity at shorter timescales have less influence on the network than nodes with activity at longer timescales. There are two ways in which this has been meas-

---

[5] Note that there are methodological issues with identifying functional hubs based on degree alone. In Pearson correlation networks, degree is partially driven by the size and not only the amount of influence a subnetwork has. Thus, nodes in larger brain areas tend to be identified as hubs in because they are part of large physical entities (Power et al. 2013). Yet some areas consistently come out as hubs in functional connectivity studies using different measures, such as anterior and posterior cingulate gyrus of the DMN (van den Heuvel and Sporns 2013).

ured, one comparing temporal activity between areas in response to a given event, and another focusing on the oscillatory properties of brain areas.

In a measure of the first type, Deco and Kringelbach (2017) determined the *integration value* of a node's activity in response to an event—for instance the presentation of a stimulus. A node's integration value is given by the number of other nodes to which it is functionally connected after the event. The higher the integration value, the higher is its influence on the network during the time period in question. This can be extended to changes in overall functional states of the brain, such as the change from wakefulness to sleep, or the induction of a coma.

Deco and Kringelbach's computational modeling of the distribution of integration values suggests that the brain is organized into a graded, non-uniform hierarchy. There exists a continuum between nodes with a small and local influence and nodes with a large and global influence on the network. Only few nodes are situated at the top of this hierarchy, because they have large integration values and respond flexibly to neural events. Although Deco and Kringelbach do not report where these nodes are located in the brain, their modeling results mirror other functional connectivity studies which report a graded hierarchy (Margulies et al. 2016), with few hub nodes at the top (Meunier et al. 2009).

The second way of applying the topological approach to the temporal domain involves *oscillatory* hierarchies (Lakatos et al., 2005). Background activity within a brain area, often known as a *local field potential*, oscillates at characteristic frequencies. It is a widespread finding that lower-frequency oscillations constrain or modulate activity at higher frequencies and spiking behavior, either via phase coupling or phase-amplitude coupling (Canolty & Knight, 2010). Moreover, *synchrony* in oscillatory phase between distinct brain areas, especially at lower frequencies, is often posited to be a key principle underlying neuronal communication and functional cooperation, and these principles have been posited to underlie recruitment of task-specific networks (Canolty et al., 2010). Intriguingly, different oscillatory frequencies have different distributions in the brain,

and low-frequency oscillations are highly exhibited in hubs which overlap with the DMN (De Domenico, Sasai, & Arenas, 2016). Thus, oscillatory hierarchies are one way in which network centrality can integrate information across the brain (cf. Burnston, forthcoming).

The above examples show that researchers using a topological approach understand hierarchical position as the amount of influence a node has on the network, either by anatomically connecting many other nodes in space (centrality) or by functionally connecting them in time (integration value or phase synchrony). This focus on network influence makes topological approaches *neutral* with regard to the representational architecture of the brain. Although many studies we describe in this section do interpret their results functionally, the assumptions from which these interpretations are derived are not part of the graph-theoretic measures themselves (see section 4.2 below). A graph-theoretic description of a node simply characterizes and quantifies its relationships to other nodes. It does not determine what information is exchanged via these connections or how.

Some researchers make this neutrality explicit: "our goal was not to identify unique hierarchical arrangements of brain regions, in terms of representational stages of streams, an approach taken in earlier work" (da Costa & Sporns 2005, p. 573; "earlier work" refers to studies following the representational approach). Instead of determining which perceptual features are represented at each level of the visual representational hierarchy, da Costa and Sporns analyzed how each node spreads its outgoing connections throughout the network hierarchy, defined in terms of degree distance. Similarly, topological methods can detect modules in a "purely data-driven way" (Sporns and Betzel 2016, p. 19.3), without using prior knowledge about the representational function of brain systems to detect modular community boundaries. Because they are neutral about representational function, topological approaches are also not committed to the claim that more abstract representations are processed at higher "levels" of the hierarchy. A high-degree node can be central regardless of whether it spreads modality-specific or multimodal information throughout the network. Hubs can be detected by their centrality

measurements without assigning degrees of abstraction to what they may represent.

Dynamic measurements of topological hierarchy are similarly neutral about representational architecture. For example: intrinsic ignition capability is defined by a node's integration value, i.e. the degree of broadcasting information in the network, not the type of information a node represents (Deco and Kringelbach 2017). In sum, novel topological approaches to hierarchy focus on the influence and the spatiotemporal propagation structure of signals and are neutral with regard to the representational function at different levels of neural hierarchy. This very neutrality is what allows for the variety of possible relationships one might posit between the representational and topological hierarchy. We move to discuss those relationships in the next section.

4. The relationship between the representational and topological views

4.1. Stage setting

Neuroscientists using graph-theory are often unclear about the precise relationship between representational and topological approaches. Sporns (2011) sometimes seems to suggest that both approaches can be combined. He claims that network structure in the brain reveals that neural function is both "integrated" and "segregated". Segregation involves the separation of the network into distinct functional units, and integration involves the exchange of information between those units. However, Sporns also writes that the topological hierarchy presents a challenge to the representational view: "Even cursory examination of structural brain connectivity reveals that the basic plan is incompatible with a model based on predominantly feedforward processing within a uniquely specified serial hierarchy" (Sporns, 2011, p. 150). How should we interpret these opposing tendencies?

We suggest construing the situation as follows. The concept of hierarchy is currently a patchwork, consisting of two approaches to hierarchical relations between brain parts. The representational approach provides researchers with particular explanatory schemas, which interpret hierarchical levels based on how abstract the representations they process are, and the input-output relations between them. The topological approach provides researchers with graph theoretical concepts like topological centrality or temporal integration to infer hierarchical levels based on a node's influence on the network. It is, however, currently an open question how these different connotations of "hierarchy" are *related* to one another.

In the following we discuss three possible relationships. On the one hand the fact that network models could explain how functions can be differentiated and how information can flow between them might suggest that "presumed aspects of the sequential organization of brain networks can be confirmed and clarified through formal topological analysis" (Hilgetag and Goulas 2020, 5). We call this the *substantiation view*. On the other hand, the high degree of interactivity in networks suggests that clear hierarchical orderings in the processing of information may not be feasible. If this is the case, then network models may offer up *alternative* organizing principles for the brain, based around the topological notion of hierarchy, which will displace the more traditional representational view. We call this the *conflict view* of the relationship. Finally, a *pluralist view* would take both motivations into account and state that there are multiple distinct hierarchical organizations instantiated in the brain. Some situations may involve modeling it as a representational hierarchy, and some a topological one, where these neither conflict nor entirely overlap.

In what follows, we discuss the commitments of each view of the relationship, and the evidential standing of those commitments. Importantly, we note examples of individual scientists who adopt, without conceptual argument, one kind of view or another. This shows that scientists themselves are being guided by particular semantic intuitions about the notion of hierarchy. The analysis thus exposes both the current state of the concept of hierarchy and articu-

lates the argumentative burden of different approaches to its patchwork structure.

4.2. The substantiation view

The substantiation view holds that the representational and topological approaches, despite using different methods, measure the *same* hierarchical organization in the brain, although the latter perhaps with a more detailed understanding of connectivity. The perspectives, after all, draw from overlapping datasets. The CocoMac database, for instance, is a database of anatomical connections based on histological data. It is frequently used for analyses within the topological approach, but includes the data that Felleman and Van Essen used to model the representational hierarchy. Two further motivations for the substantiation view are (i) that the modularity of networks can be interpreted as underlying distinct functions of the type posited in the classical hierarchy, and (ii) that the topological divisions revealed through network analysis often match functional divisions posited by the representational approach. We will discuss these briefly in turn.

First, point (i). Recall that, on the representational view, each neural system (visual, motor, frontal, etc.) exhibits significant functional autonomy from other systems. Further, within each system, the distinct areas play different functional roles in performing the system's overall function. One possible way of reading the modular architecture of topological hierarchies is as implementing functionally specified subsystems, whose integration then proceeds in, at least roughly, the way described by the representational view. Modules, recall, are characterized as parts of the network with primarily intra-module connections, thus supporting the notion that they are computational units dedicated to specific kinds of problems. Indeed, Meunier, Lambiotte, & Bullmore (2010) suggest that a hierarchy of modules allows for each module to "specialize in sub-problems." Breakspear and Stam (2005) argue that lower levels of the topological hierarchy "represent specific features." (To be fair, both papers note that integrating information from distinct modules may be a global process.) The conceptual possibility of topological modules underlying the

specific functions and interactions posited in the representational view is alluring to those friendly to the representational approach.

The support for point (ii) is empirical. It turns out that, in fact, many divisions made within the topological approach correspond to divisions made within the representational approach. This is especially true for large-scale divisions (but see Zerilli, 2017). For instance, modularity analyses at the level of the whole brain reveal that visual cortex is more tightly interconnected than it is connected to other large-scale networks. In cats and macaques visual cortex is much more tightly interconnected than it is connected to somatosensory cortex, and vice versa (Sporns, Honey, & Kötter, 2007). This is true for both structural and functional connections (Honey, Kötter, Breakspear, & Sporns, 2007). Even within these parts of the cortex, functional divisions can be made that match the representational view – for instance, structural connectivity in humans shows a distinction between the dorsal and ventral streams of the visual cortex (Hagmann et al., 2008), which are standardly taken to perform very different functions in vision (Mishkin, Ungerleider, & Macko, 1983).

Moreover, areas of cortex that have traditionally been called "association areas," including areas in the parietal and prefrontal cortices, standardly come out as hubs in graph-theoretic network analyses (Sporns et al., 2007, van den Heuvel & Sporns 2013). If their role is to associate (and perhaps abstract from) multiple kinds of information from unimodal cortices, then one would expect them to have a wide range of connections to those areas. Sporns (2011) himself cites approvingly the unimodal-to-association area progression posited by Mesulam and others (cf. Meyer & Damasio, 2009). Passingham, Stephan and Kötter (2002), in an influential analysis, proposed that areas such as premotor and frontal cortices will differ in the amount of different information they will respond to from sensory cortices, and that these differences are due to differences in the patterns of connections exhibited by different areas.

Researchers using resting state functional connectivity studies have also embraced the substantiation view. Margulies et al. (2016) used diffusion map embedding, a variety of dimensionality reduction

technique, on human resting state functional connectivity data. This technique involved constructing dimensions along which connected areas could be grouped, with closely connected areas close together along each dimension. The sum of all dimensions forms a so-called *embedding space*, which positions nodes according to the similarity of their functional connectivity profiles. In Fig. 5. Margulies et al. use two of these dimensions to describe the greatest and second greatest amount of variance in functional connectivity between areas, which they call the first and second *gradient* of connectivity.
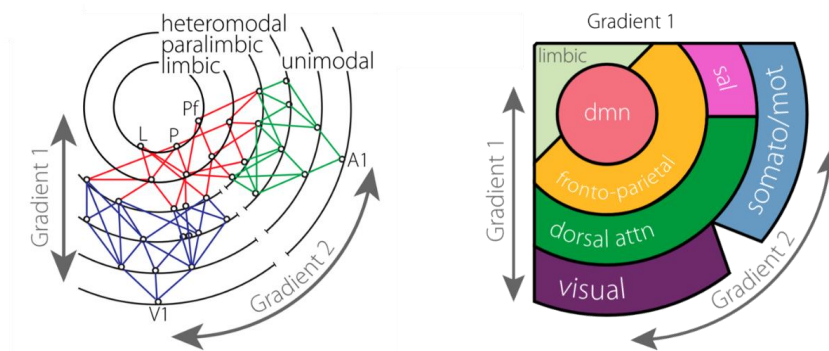

Fig. 5. The Mesulam model (left) and the Margulies model (right) of the cortical abstraction hierarchy. Adapted from Margulies et al. (2016).

Fig. 5 shows that Margulies et al. interpret the two gradients of functional connectivity as revealing a *hierarchical gradient of abstraction* which runs from primary sensory areas to regions of the default mode network (DMN). According to this interpretation, default mode regions are involved in cognitive functions such semantic memory or reward-guided decision making because default mode activity processes abstract informational content, largely independent of transient environmental stimuli processed by sensory systems.

This interpretation substantiates Mesulam's representational hierarchy model (see section 2) because it situates the DMN at the top of a known representational hierarchy that proceeds from unimodal sensory to transmodal association areas. Note, however, that this substantiation interpretation is not necessary to apply the diffusion map embedding algorithm to resting state fMRI data. This procedure places nodes closer in embedding space if they are more strongly functionally connected, or as we put it, if they influence each other

more strongly than other nodes. Additional assumptions about functional connectivity directly reflecting information processing (Schölvinck et al. 2013) and topographical structure constraining cognitive processes are required (Margulies et al. 2016). To arrive at the substantiation view, these assumptions need to be combined with the supposition that the representational approach is a correct approximation of the brain's hierarchical organization.

4.3. The conflict view

There are two primary motivations for the conflict view: (i) graph-theoretical results that conflict with the representational hierarchy; and (ii) independent evidence that speaks against the representational but not the topological approach. We take these motivations in turn.

There are individual cases in which the consistency between topological and representational approaches to hierarchy breaks down. Let us consider one case – V4 – in detail. V4 is, according to the representational approach, a "mid-level" visual area (level 5 of Felleman and Van Essen's hierarchy), which comprises two sub-areas in charge of representing color and complex shape. This clear place in the representational hierarchy is questioned by graph theoretic analyses of anatomical connectivity, which reveal that V4 scores extremely highly in measures of degree and centrality. This is shown in figure 6 below.
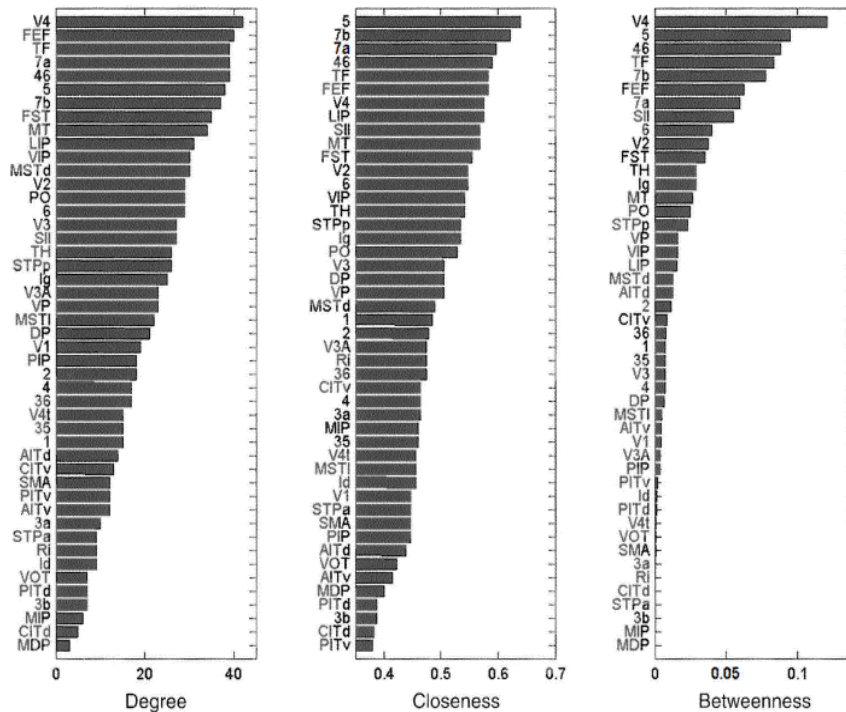
Fig. 6. Centrality measurements of V4. From Sporns (2011).

Fig. 6 shows that V4 scores very highly, relative to the whole-brain network, on degree and betweenness centrality. It also ranks high on closeness centrality, which is a related measure of the average path length between the node and all other nodes in the network (shown in the inverse here for comparative ranking). V4 *also* has connections to other high centrality nodes, such as area 46 in the frontal cortex. Similarly, nodes that are directly connected with V4 (da Costa and Sporns' hierarchical level 1), have a low clustering coefficient, but nodes that are connected to *those* nodes (da Costa and Sporns' hierarchical level 2) have a very high clustering coefficient. This suggests that V4 connects, with a small number of synaptic steps, to multiple modular areas (da Costa & Sporns 2005). For areas in the dorsal stream such as MT and MST, by contrast, clustering is greater at nodes only one edge away. The way to interpret this is that most connections for dorsal stream areas are intra-modular, whereas connections for V4 are widely spread across mod-

ules. Thus, V4 is potentially a more integrative area than areas that are traditionally posited to be at the same or higher levels of the representational hierarchy.

This result suggests that, in terms of topological centrality, V4 is at the *highest* levels of the *overall* brain hierarchy, in extreme contradistinction to the low level posited for it in the representational hierarchy. Hence, there is a direct conflict between the results within the two different perspectives. Does the centrality of V4 make a functional difference? As Sporns notes, hubs are well-situated to play multiple diverse functional roles, and this is in fact borne out by the data – V4 has a much more complex functional profile than the representational hierarchy posits (Burnston, 2016b; Roe et al., 2012), and lesions to V4 cause a diverse range of effects (Schiller, 1993). This puts pressure on the representational view in two ways. First, V4 may not have a well-defined place in a representational hierarchy, such that it sends a specific signal onwards to subsequent areas of the hierarchy. Second, it pressures the idea that sensory representational occurs *first,* prior to the integration of multimodal information by association areas.

The second motivation for the conflict view is independent anatomical and physiological data that conflict with the functional posits of the representational hierarchy. We can only summarize this data here, but it will suffice to get the picture across. First, both direct and subcortically mediated connections exist between primary sensory cortices in different modalities, and these are posited to underlie a variety of cross-modal effects (Driver & Spence, 2000, Ghazanfar and Schroeder 2006). Second, the representational approach suggests a preferred pathway for signals in sensory cortices, such that information is represented first at lower levels, then only subsequently at higher levels (Lamme & Roelfsema, 2000). However, both anatomical and time course data question the existence of such a pathway. Parts of V4 have both bidirectional and direct connections to higher visual areas which bypass the putative central ventral pathway, "violating a strict serial hierarchy at even the earliest stages of visual processing" (Kravitz et al. 2013). Temporal data show V4 in fact is *slower* to represent information than areas traditionally seen as "above" it in the hierarchy such as MST and the FEF,

whereas MT is roughly tied with these areas in terms of response latency. This result is summarized in Fig. 7 below.
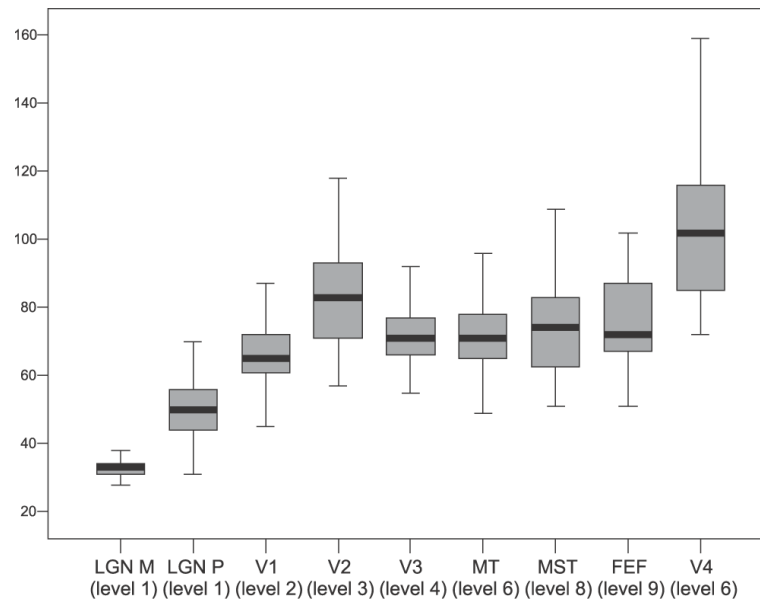

Fig. 7. Time-from-stimulus onset measurements for physiological activation of visual cortical areas. From Capalbo et al. (2008). "Level" refers to hierarchical level, in the sense of Felleman and Van Essen (1991), except Capalbo et al begin counting from the LGN, rather than V1. Hence, e.g., MT and V4 are labelled as "level 6" here, but they are level 5 in Felleman and Van Essen.

Third, physiological results question the idea that increasingly abstract representation occurs at higher levels. Hedge and Van Essen (2007) measured physiological responses in V1, V2, and V4 to a wide range of shapes. Examples are shown in Fig. 8 below.
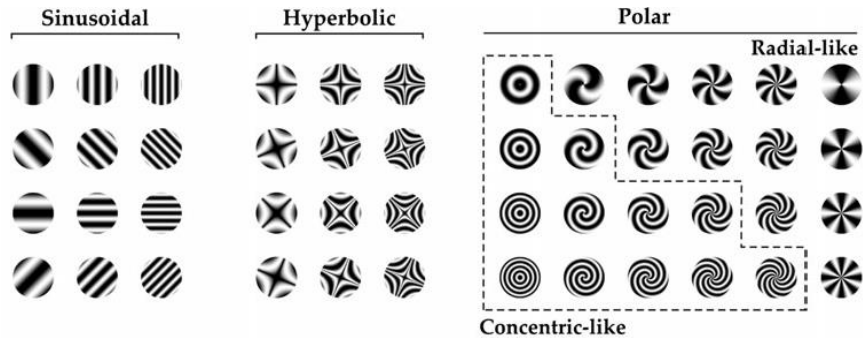
Fig. 8. Shapes of increasing complexity. From Hegde and Van Essen (2007).

According to the representational hierarchy, more complex shapes should be represented in higher areas of the hierarchy – in this example, simple sinusoidal gratings should be represented at V1 and V2, while increasingly complex hyperbolic and polar/radial shapes should be represented at V4. But this is not what Hegde and Van Essen found. Instead, they showed that different populations of cells in each area had greater responses to shapes *across the categories*, without one type of shape being privileged at any area. Strikingly, the authors – including Van Essen, one of the key progenitors of the representational hierarchy view – argue that their data undermines any strict division between what is represented at distinct representational stages in the visual cortex.

These results generalize both to relationships "higher up" in the purported processing hierarchy, and to the motor domain. For instance, Meyers et al. (2008) compared how much category information about a stimulus is extractable from populations in the inferotemporal and prefrontal cortices. There was no difference in the degree of abstraction of information that can be discerned from these populations (using decoding methods). What differed is what information co-existed with abstract category information in each population. IT tended to retain more visual detail, whereas PFC tended to combine stimulus category information with task variables. Similarly, Murray et al. (2017) modeled the circuit between prefrontal and posterior parietal cortex involved in working memory. They showed that the difference between the PFC and PPC is not how abstractly they represent information, but instead in terms of whether they *also* represent distractors – PPC does whereas PFC

does not. These results suggest that areas at different levels of the traditional hierarchy are not distinguished by how abstractly they represent information, but in terms of how they represent different combinations of information that are relevant for a task.

As a final example, consider the notion of rule-representation in the frontal cortex. Rules are often construed as related to either conditional stimulus-response associations or as generalizations of those associations. So, in a same-different task, one might have neurons that respond both to the stimulus and to its repetition, or one might have cells that signal when the task is a same-different task, *regardless* of the stimulus. The latter is generally construed as more abstract, but cells with significant responses for rule do not distribute hierarchically in the cortex. In fact, rule selectivity has been shown to occur *more strongly* and *earlier* in a task in the premotor cortex than the prefrontal cortex (Wallis & Miller, 2003). Moreover, there are not individual areas that represent rules at the expense of stimuli – in fact, the much more common finding is that cells even in areas traditionally construed as higher-level exhibit *mixed selectivity*, with responses mediated by combinations of stimulus, response, and rule (for review, see Rigotti et al., 2013).

In general, the results here, along with the results in visual cortex discussed above, do not support the notion of a clear hierarchy of representational abstraction either in visual or "association areas" – instead, what differentiates areas is how they combine different information in different ways.

One worry about the conflict view is that because topological approaches are neutral with regard to representational architecture, there is no inherent reason to align them with independent evidence against the plausibility of the representational view. The fact that topological approaches are compatible with that evidence does not entail that they positively supports that evidence. Our reply is that at least in some cases, graph-theoretical analyses *do* support evidence against the representational view, despite their neutrality towards representational function. Consider, for instance Goulas et al. (2014), who tested predictions about anatomical connectivity entailed by the anterior-posterior gradient of abstraction in the prefrontal cortex. They reasoned that, if more anterior areas of the prefron-

tal cortex were in charge of more abstract control functions, then they should send more efferent connections to areas lower in the purported hierarchy than they receive. Goulas et al. (2014) could not confirm this prediction of the abstraction gradient model, however. More posterior prefrontal regions, Brodmann areas 45 and 46, consistently sent more efferent connections than the most anterior region, area 10. Therefore, the anatomical connectivity of these regions conflicts with the anterior-posterior model.

We have shown that, despite the consistencies between the representational and topological approaches, there is also data that the topological approach can accommodate, that the representational one cannot, or at least not easily. Hence, the two views are empirically distinguishable. If one finds the data reviewed in this section compelling, one is likely to adopt the conflict view and suggest displacement of the representational approach by the topological one.

4.4. The pluralist view

Both the substantiation and the conflict view seek to resolve the patchwork structure in favor of a univocal meaning of the concept of "hierarchy", referring to a distinctive organizational property. Substantiation implies that distinct hierarchical levels must always correspond to degrees of representational abstraction, and are individuated in terms of representational function. Conflict implies that hierarchical distinctions are always specified in terms of amount of influence, and are individuated with no representational commitments.

One might reasonably suspect, however, that any attempt to build a universal conceptual structure of "hierarchy" is mistaken, given the piecemeal data upon which the substantiation and conflict views are founded. Instead, one could propose a *pluralist* view about the relation between representational and topological approaches: they represent multiple, equally legitimate meanings of "hierarchy" in neuroscience which overlap in some domains and diverge in others. Pluralism suggests that both representational hierarchy and topological approaches, while having distinct constitutive commitments, are explanatorily important for understanding neural organization. Plu-

ralists hold that the extant patchwork structure of scientific concepts is epistemically useful and—to a certain extent—reflects the structure of the underlying phenomena (Wilson 2006, Bursten 2016, Novick 2018, Haueis 2018). Below we highlight three pluralist options and discuss their advantages and drawbacks.

The first option is that there are different *processes* in the brain which will be best explained by the representational and topological approaches. On this view, there is a large amount that is correct to the representational approach – the basically serial and abstractive nature of processing, for instance – but this process breaks down at some point and gives way to a different form of organization that relies more on global interactivity. This form of pluralism is suggested by some of the comments from theorists discussed in section 4.1. The basic problem with this form of pluralism is that it does little to answer any of the data that speaks against the representational hierarchy, since it basically accepts the traditional picture and views the topological hierarchy as a kind of integrative add-on.

The second form of pluralism is a *modelling*-based pluralism, which treats the representational and topological approaches as ways of *representing* the brain. On this view, both the representational and topological approaches can be seen as strategies for understanding neural organization, where the reason for adopting one over another depends on the explanandum. Network representations can be used to think about, for instance, efficiency of communication given constraints such as minimizing wiring length (Meunier et al., 2010; van den Heuvel & Sporns, 2011). This might be contrasted with the representational hierarchy, which is meant to explain how signals are in fact processed in the brain. While this view has some advantages, and connects up with larger debates about the role of different forms of models in explanation in biology (Green et al., 2017), an explanation will have to be given about the situations in which these models conflict, such as in the case of V4 discussed above.

The other way to accommodate conflicting data is *organizational pluralism*, which suggests that the brain can in fact instantiate many different *forms* of organization, and that the representational hierar-

chy is one but not the only one. For instance, in many studies that inspire the representational approach, animals are studied in very limited behavioral circumstances, having to make specific perceptual judgments on the basis of presented stimuli (in the perceptual case), or having a well-defined task set that they must learn (in the prefrontal case). Perhaps, however, perception *in the context of action* requires more dynamic interaction with wider brain networks, or action in the case of *deliberation* requires broader access to, e.g., motivational and evaluative influences. On this view, there is a simple hierarchical organization for simple behavioral contexts, but this organization might be replaced by more complicated forms of signal representational, which might also be mediated by the topological hierarchy (cf. Chemero & Silberstein, 2013).

We think the last view is in many ways the most promising, although not without limitations. One advantage of organizational pluralism is that it comports with a wide range of data suggesting that the network organization of the brain is not *constant* (Honey et al., 2007). When analyzing functional connectivity, different nodes attain different degrees of centrality in different contexts, and different networks are enlisted that are relevant to the task (Burnston, 2019; Stanley, Gessel, & de Brigard, 2019). Organizational pluralism accounts for this possibility while making room for the traditional representational picture as one kind of organization that the network can adopt. Another advantage is that organizational pluralism can in principle account for *both* the data in favor of, and the data against, the representational hierarchy view. If the organization of the brain changes dynamically, then in some cases it might instantiate a representational hierarchy, while in some cases it may not – hence the traditional data in favor of, as well as the newer data against, the representational view.

The main worry about this last view is that it may be too permissive. For instance, the latency data from Capalbo et al., as well as the physiological data from Hegde and Van Essen, seem to cause problems for the representational view *even in the kind of contexts* for which it was originally proposed. Organizational pluralists must be able to account for data in the same contexts via proposed changes in organization.

In sum, we suggest that the substantiation, conflict and pluralist views are all both independently motivated (to some degree) and at work in the current literature. Given that they are all *distinct* views, however, they need to be articulated, and their commitments understood, in order for conceptual progress to be made. We have offered a preliminary version of such a framework above. In the next section, we showcase the utility of this framework by applying it to recent research on brain dynamics and rich club topology.

5. A test case: rich club organization

As research into brain networks has progressed, attention has turned heavily towards brain dynamics, and how they are shaped by network features, including hierarchical centrality. Earlier, we discussed how the hub-and-module organization of the brain is often seen as a way of implementing the balance between segregation and integration of function. A dynamical corollary to this view is that highly central nodes allow for a balance of diffusion and efficiency – diffusion means that information can be broadcast widely in the network, while efficiency means that it can be routed to where it is needed (Avena-Koenigsberger, Misic, & Sporns, 2017). Whole-brain dynamics shift between rest and task, and between tasks (Shine & Poldrack, 2017), and are mediated by widespread oscillatory synchronization (Deco & Kringelbach, 2016).

In this section, we briefly discuss the role of the "rich-club" architecture in the brain for mediating dynamics. A network contains a rich club if its highest-degree nodes are also highly connected to each other. A rich club measurement begins with a degree threshold, k, and then asks what proportion of possible connections between nodes with degree > k obtains in the network. Rich club architectures occur in many networks, including the human brain. Simulations have shown that brain networks with a rich-club architecture have a greater range of dynamic attractors than networks without one (Senden, et al., 2014).

Rich-club architecture provides an interesting test case for the different positions relating representational and topological hierarchies.

First, rich club areas are at the highest levels of network-based centrality, judged by degree and influence on cortical dynamics. It also operates at particularly slow oscillatory frequencies (Senden et al. 2017a), placing it at a higher level of the oscillatory hierarchy. Finally, the rich club also significantly overlaps with the DMN which, as we saw above, is posited within the substantiation view to be transmodal network with the highest degree of abstraction in the brain (Margulies et al. 2016). So, the rich club is a particularly rich network structure with which to analyze concepts of hierarchy.

The representational hierarchy view posits that the rich club should be involved in processing abstract representations. However, an alternative hypothesis has emerged from within the network literature. Senden et al. (2017b) studied functional connectivity between the rich club and other brain areas as subjects switched between rest and four different kinds of tasks, including working memory (n-back), response inhibition, mental rotation, and verbal reasoning. They showed, intriguingly, that during rest the rich club network had greater in-degree, meaning it received more input from other brain areas, but that this switched during tasks, with the rich club providing more output than receiving input. Moreover, rich club outputs targeted a similar set of brain areas across tasks, but the network relationship between these target areas, as well as which brain areas *they* interacted with, changed depending on the task.

The hypothesis constructed by Senden et al. is that the rich club serves as a *gate* that *mediates competition* between networks elsewhere in the brain that mediate the specific tasks. Note that, as befitting the different core commitments of the topological view, the gating hypothesis contains no commitments about whether the rich club does this by conveying abstract representations about task context to the rest of the network, or even whether it represents anything at all. Hence, all of the positions with regards to the relationship between the two views of hierarchy are on the table. We will not attempt to adjudicate which is correct here, but we will close by listing the explanatory obligations that each view of the relationship takes on.

The substantiation view suggests that the rich club's influence on the rest of the network is dependent on the rich club processing par-

ticularly abstract representations. A proponent of the substantiation view must then define *what* those representations are, how they are propagated to the non-rich-club nodes that receive input from the rich club, and how these are used to guide behavior. A proponent of the conflict view must argue that the competition-process is mediated primarily by the slow-oscillation and conflicting inputs about task settings coming into the rich club, and that no abstract representations are required to subsequently re-organize its output targets in the appropriate configuration for the task.

Each variety of pluralist view is also possible. *Process* pluralists would suggest that input about the task settings, and perhaps motor actions involved in implementing particular tasks, follow a representational hierarchy, but that coordinating the different subnetworks is itself a topological, and not a representational hierarchy-based, process. Model-based pluralism will suggest that the roles of the rich club in mediating diffuse yet efficient communication, as well as providing robust communication (van den Heuvel & Sporns, 2011) are best described from the topological perspective, but that this is compatible with abstract representations being what is communicated diffusely and efficiently. Finally, organizational pluralism states that rich club organization, which is topological, co-exists with representational hierarchies in the brain, perhaps explaining why in-degree is significantly higher to the rich-club between tasks, but out-degree higher when task-related representations are occurring.

Each of these views in turn takes on commitments, particularly with regards to how the other areas with connections to the rich club operate. The point is that *none* of these moves is trivial, and hence whatever position one takes requires extensive justification. So, our approach to the patchwork concept helps clarify the state of the hierarchy concept with regards to extant research strategies and the available empirical data.

6. Conclusion

In this paper we have argued that there are two distinct approaches to the concept of hierarchy in neuroscience, whose relations have not been sufficiently scrutinized in the previous literature. While the representational approach takes progressively more abstract information processing and representational function as the core property which sorts anatomical areas hierarchically (section 2), topological approaches take influence on the network and propagation structure to be central and are neutral with regard to abstraction and representational function (section 3).

Our analysis of these two approaches supports the descriptive claim that many scientific concepts develop into a patchwork when researchers use them to pursue various descriptive and explanatory projects (Wilson, 2006; Bursten 2016; Novick, 2018; Haueis, 2018). Our central contribution is the point that such conceptual patchworks leave researchers with multiple options of how to relate different uses of a concept to each other. We argued that current evidence suggests three possible conceptual relations between the two approaches to "hierarchy" (section 4): topological hierarchies could *substantiate* the traditional representational hierarchy, *conflict* with it, or contribute to a *plurality* of approaches needed to understand the hierarchical organization of the brain. We do not wish to argue which of these relations is the correct one. We take the foregoing to have shown, however, that the conceptual landscape surrounding the notion of "hierarchy" in systems neuroscience is *extremely* complicated. Without explicating its different connotations and their relations, "use of the term 'hierarchy' can become meaningless, or worse, misleading" (Hilgetag and Goulas 2020, 8). There are no obvious answers, and there is especially no justification to presuming one view of the relationships between different notions of hierarchy over another.

Because hierarchical thinking is deeply engrained in neuroscience and is also used to defend computational (Pylyshyn, 2007) and evolutionary (Barrett, 2014) accounts of the mind, theorizing about relationship between the representational and topological views is of no small consequence for cognitive science. A substantiation view allows for standard conceptions of the general architecture of the brain

and mind to be kept in place, with perhaps some network concepts used to fill in details or account for information integration in a more perspicuous way. The conflict view, however, promotes – and we want to stress this – a *radical* revision to our general conception of neural and mental organization, for which there are not well-articulated alternatives. Thinking about the representational and functional organization of the brain if the conflict view is true is a major conceptual project. Finally, if one pursues a pluralist option then examining the nature of the interaction between different notions of hierarchy will generate insight about functional architecture and the roles of distinct concepts in neuroscience. By articulating different possibilities of answering that question, we hope to have opened up a conceptual space in which further neuroscientific and philosophical reasoning about neural hierarchy can proceed.

References

Avena-Koenigsberger, A., Misic, B., & Sporns, O. (2017). Communication dynamics in complex brain networks. *Nature Reviews Neuroscience, 19*(1), 17-33. doi:10.1038/nrn.2017.149

Badre, D., Kayser, A. S., & D'Esposito, M. (2010). Frontal cortex and the discovery of abstract action rules. *Neuron, 66*(2), 315-326.

Barrett, H. C. (2014). *The shape of thought: How mental adaptations evolve*: Oxford University Press.

Bastos, A.M., Usrey, W.M., Adams R.A., Mangun G.R., Fries P. & Friston K.J. (2012). Canonical microcircuits for predictive coding. *Neuron 76*, 695–71

Bechtel, W. (2008). *Mental mechanisms: Philosophical perspectives on cognitive neuroscience*. New York: Routledge.

Breakspear, M., & Stam, C. J. (2005). Dynamics of a neural system with a multiscale architecture. *Philosophical Transactions of the Royal Society of London. Series B, Biological sciences, 360*(1457), 1051-1074.

Burnston, D. C. (2016a). Computational neuroscience and localized neural function. *Synthese, 193*(12), 3741-3762.

Burnston, D. C. (2016b). A contextualist approach to functional localization in the brain. *Biology & Philosophy, 31*(4), 527-550.

Burnston, D. C. (2019). Getting over atomism: Functional decomposition in complex neural systems. *British Journal for the Philosophy of Science*. doi:10.1093/bjps/axz039

Bursten, J. (2016). Smaller than a breadbox: Scale and natural kinds. *British Journal for the Philosophy of Science, 69*(1), 1-23.

Canolty, R. T., Ganguly, K., Kennerley, S. W., Cadieu, C. F., Koepsell, K., Wallis, J. D., & Carmena, J. M. (2010). Oscillatory phase coupling coordinates anatomically dispersed functional cell assemblies. *Proceedings of the National Academy of Sciences, 107*(40), 17356-17361.

Canolty, R. T., & Knight, R. T. (2010). The functional role of cross-frequency coupling. *Trends in Cognitive Sciences, 14*(11), 506-515.

Capalbo, M., Postma, E., & Goebel, R. (2008). Combining structural connectivity and response latencies to model the structure of the visual system. *PLoS Computational Biology, 4*(8), e1000159.

Craver, C.F. (2007). *Explaining the brain. Mechanistic explanation and the mosaic unity of neuroscience.* Oxford: Oxford University Press.

Deco, G., & Kringelbach, M. L. (2016). Metastability and coherence: Extending the communication through coherence hypothesis using a whole-brain computational Perspective. *Trends in Neurosciences, 39*(3), 125-135. doi:10.1016/j.tins.2016.01.001

Deco, G. & Kringelbach, M.L. (2017). Hierarchy of information representational in the brain: A novel 'intrinsic ignition' framework. *Neuron 94*, 961–968.

da F. Costa, L., & Sporns, O. (2005). Hierarchical features of large-scale cortical connectivity. *The European Physical Journal B, 48*(4), 567-573.

De Domenico, M., Sasai, S., & Arenas, A. (2016). Mapping multiplex hubs in human functional brain networks.

Frontiers in Neuroscience, 10, 326. doi:10.3389/fnins.2016.00326

Driver, J., & Spence, C. (2000). Multisensory perception: Beyond modularity and convergence. *Current Biology, 10*(20), R731-R735.

Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical representational in the primate cerebral cortex. *Cerebral cortex, 1*(1), 1-47.

Ghazanfar A.A., Schroeder C.E. (2006). Is neocortex essentially multisensory? *Trends in Cognitive Science 10*(6), 278–285.

Grafton, S. T., & de C Hamilton, A. F. (2007). Evidence for a distributed hierarchy of action representation in the brain. *Human Movement Science, 26*(4), 590-616.

Green, S., Şerban, M., Scholl, R., Jones, N., Brigandt, I., & Bechtel, W. (2017). Network analyses in systems biology: New strategies for dealing with biological complexity. *Synthese* 195(4), 1751–1777.

Gross, C. G., Rocha-Miranda, C., & Bender, D. (1972). Visual properties of neurons in inferotemporal cortex of the Macaque. *Journal of Neurophysiology, 35*(1), 96-111.

Goulas, A., Uylings, H.B.M. and Stiers, P. (2014). Mapping the hierarchical layout of the structural network of the macaque prefrontal cortex. *Cerebral Cortex 24*, 1178–1194.

Guimerà, R., & Amaral, L. a. N. (2005). Cartography of complex networks: Modules and universal roles. *Journal of Statistical Mechanics, 2005*(02), P02001.

Hagmann, P., Cammoun, L., Gigandet, X., Meuli, R., Honey, C. J., Wedeen, V. J., & Sporns, O. (2008). Mapping the structural core of human cerebral cortex. *PLoS biology, 6*(7), e159-e159. doi:10.1371/journal.pbio.0060159

Haueis, P. (2012). The fuzzy brain: Vagueness and mapping connectivity in the human cerebral cortex. *Frontiers in Neuroanatomy 6* (37), doi: 10.3389/fnana.2012.00037.

Haueis, P. (2018). Beyond cognitive myopia: A patchwork approach to the concept of neural function. *Synthese 195*(12), 5373–5402. doi: 10.1007/s11229-018-01991-z.

Hegdé, J., & Van Essen, D. C. (2007). A comparative study of shape representation in macaque visual areas V2 and V4. *Cerebral cortex, 17*(5), 1100-1116. doi:10.1093/cercor/bhl020

Hilgetag C.C., O'Neill M. & Young M.P. (1996). Indeterminate organization of the visual system. *Science 271*(5250), pp. 776–777.

Hilgetag C.C., & Goulas, A. (2020). 'Hierarchy' in the organization of brain networks. *Philosophical Transactions of the Royal Society B* 375: 20190319. doi:10.1098/rstb.2019.0319.

Hohwy, J. (2013). *The predictive mind*: Oxford University Press.

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 of the United States of America, 104*(24), 10240-10245.

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of Physiology, 160*(1), 106.

Hütt, M.-T., Kaiser, M., & Hilgetag, C. C. (2014). Perspective: Network-guided pattern formation of neural dynamics. *Philosophical Transactions of the Royal Society B: Biological Sciences, 369*(1653), 20130522.

Huneman, P. (2010). Topological explanations and robustness in biological sciences. *Synthese 177*, 213– 245

Koechlin, E., Ody, C., & Kouneiher, F. (2003). The architecture of cognitive control in the human prefrontal cortex. *Science, 302*(5648), 1181-1185.

Kostić (2016). The topological realization. *Synthese, 195*(1), 79-98. doi 10.1007/s11229-016-1248-0

Kravitz, D.J., Saleem, K.S., Baker, C.I., Ungerleider, L.G. and Mishkin, M. (2013). The ventral visual pathway: an expanded neural framework for the processing of object quality.

Lakatos, P., Shah, A. S., Knuth, K. H., Ulbert, I., Karmos, G., & Schroeder, C. E. (2005). An oscillatory hierarchy controlling neuronal excitability and stimulus representational in the auditory cortex. *Journal of Neurophysiology, 94*(3), 1904-1911.

Lamme, V. A., & Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent representational. *Trends in Neurosciences, 23*(11), 571-579.

Margulies, D.S. Ghosh, S.S., Goulas, A., Falkiewiz, M., Huntenburg, J.M., Langs, M., Bezgin, G., Eickhoff, S.B., Castellanos, F.X., Petrides, M., Jefferies, E. and Smallwood, J. (2016). Situating the default mode network along a gradient of macroscale cortical organization. *PNAS 113*(44), 12574–79.

Mesulam, M. (1998). From sensation to cognition. *Brain, 121*(6), 1013-1052.

Meyers, E. M., Freedman, D. J., Kreiman, G., Miller, E. K., & Poggio, T. (2008). Dynamic population coding of category information in inferior temporal and prefrontal cortex. *Journal of Neurophysiology, 100*(3), 1407-1419.

Meunier, D., Lambiotte, R., & Bullmore, E. T. (2009). Hierarchical modularity in human brain functional networks. *Frontiers Neuroinformatics 3*(37). doi: 10.3389/neuro.11.037.2009

Meyer, K., & Damasio, A. (2009). Convergence and divergence in a neural architecture for recognition and memory. *Trends in Neurosciences, 32*(7), 376-382.

Mishkin, M., Ungerleider, L. G., & Macko, K. A. (1983). Object vision and spatial vision: Two cortical pathways. *Trends in Neurosciences, 6*, 414-417.

Mushiake, H., Saito, N., Sakamoto, K., Itoyama, Y., & Tanji, J. (2006). Activity in the lateral prefrontal cortex reflects multiple steps of future events in action plans. *Neuron, 50*(4), 631-641.

Müller-Linow, M., Hilgetag, C.C., Hütt, M.-T. (2008). Organization of excitable dynamics in hierarchical biological networks. *PLoS One* 4(9): e1000190. doi:10.1371/journal.pcbi.1000190

Murray, J.D., Bernacchia, A. Freedman, D.J. Romo, R. Wallis, J.D., Cai, X. Padoa-Schioppa; Pasternak, T., Seo, H. Lee, D., Wang, X.-J. (2014). A hierarchy of intrinsic timescales across primate cortex. *Nature Neuroscience, 17*, 1661-1663. doi:10.1038/nn.3862

Murray, J. D., Jaramillo, J., & Wang, X. J. (2017). Working Memory and Decision-Making in a Frontoparietal Circuit Model. *J Neurosci, 37*(50), 12167-12186.

Novick, A. (2018). The fine structure of 'homology'. *Biology & Philosophy 33*(6). https://doi.org/10.1007/s10539-018-9617-3.

Orlandi, N. (2010). Are sensory properties represented in perceptual experience? *Philosophical Psychology, 23*(6), 721-740.

Passingham, R. E., Stephan, K. E., & Kötter, R. (2002). The anatomical basis of functional localization in the cortex. *Nature Reviews Neuroscience, 3*(8), 606-616.

Pylyshyn, Z. W. (2007). *Things and places: How the mind connects with the world*, Cambridge, MA: MIT Press.

Power, J., Schlaggar, B.L., Lessov-Shlaggar, C.N., & Petersen, S.E. (2013). Evidence for hubs in human functional brain networks. *Neuron 79*(4), 798-813.

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

Rigotti, M., Barak, O., Warden, M. R., Wang, X.-J., Daw, N. D., Miller, E. K., & Fusi, S. (2013). The importance of mixed selectivity in complex cognitive tasks. *Nature, 497*(7451), 585-590.

Roe, A. W., Chelazzi, L., Connor, C. E., Conway, B. R., Fujita, I., Gallant, J. L., . . . Vanduffel, W. (2012). Toward a unified theory of visual area V4. *Neuron, 74*(1), 12-29.

Savic, I., Gulyas, B., Larsson, M., & Roland, P. (2000). Olfactory functions are mediated by parallel and hierarchical representational. *Neuron, 26*(3), 735-745.

Schiller, P. (1993). The effects of V4 and middle temporal (MT) area lesions on visual performance in the rhesus monkey. *Visual Neuroscience 10*(4), 717-746

Schölvinck M.L., Leopold, D.A., Brookes, M.J. & Khader, P.H. (2013). The contribution of electrophysiology to functional connectivity mapping. *NeuroImage 80*, 297–306.

Senden, M., Deco, G., de Reus, M. A., Goebel, R., & van den Heuvel, M. P. (2014). Rich club organization supports a diverse set of functional network configurations. *NeuroImage, 96*, 174-182.

Senden, M., Reuter, M., van den Heuvel M.P., Goebel, R., Deco, G., (2017a). Rich club regions can organize state-dependent functional network organization by engaging in oscillatory behavior. *NeuroImage 146*, 561-574

Senden, M., Reuter, M., van den Heuvel M.P., Goebel, R., Deco, G., & Gilson, M. (2017b). Task-related effective connectivity

reveals that the cortical rich club gates cortex-wide communication. *Human Brain Mapping. 39*(3), 1246-1262.

Sennet, Adam. 2016. Polysemy. Oxford Handbooks Online, Doi: 10.1093/oxfordhb/9780199935314.013.3

Shine, J. M., & Poldrack, R. A. (2017). Principles of dynamic network reconfiguration across diverse brain states. *NeuroImage, 180*(B), 396-405. doi: 10.1016/j.neuroimage.2017.08.010.

Silberstein, M., & Chemero, A. (2013). Constraints on localization and decomposition as explanatory strategies in the biological sciences. *Philosophy of Science, 80*(5), 958-970.

Sporns, O. & Betzel R. (2016). Modular brain networks. *Annual Review of Psychology 4*(67), 613–640.

Sporns, O. (2011). *Networks of the Brain*. Cambridge, MA: MIT Press.

Sporns, O., Honey, C. J., & Kötter, R. (2007). Identification and classification of hubs in brain networks. *PloS One, 2*(10), e1049-e1049.

Stanley, M. L., Gessell, B., & De Brigard, F. (2019). Network modularity as a foundation for neural reuse. *Philosophy of Science, 86*(1), 23-46.

Uithol, S., Burnston, D. C., & Haselager, P. (2014). Why we may not find intentions in the brain. *Neuropsychologia, 56*, 129-139. doi:10.1016/j.neuropsychologia.2014.01.010

Ullman, S. (2007). Object recognition and segmentation by a fragment-based hierarchy. *Trends in Cognitive Science, 11*(2), 58-64. doi:10.1016/j.tics.2006.11.009.

van den Heuvel, M. P., & Sporns, O. (2011). Rich-club organization of the human connectome. *Journal of Neuroscience, 31*(44), 15775-15786. doi:10.1523/JNEUROSCI.3539-11.2011

van den Heuvel, M.P. & Sporns, O. (2013). Network hubs in the human brain. *Trends in Cogntive Science 17*(12), 683–696.

Vetter, P., & Newen, A. (2014). Varieties of cognitive penetration in visual perception. *Consciousness and Cognition, 27*, 62-75.

Wallis, J. D., & Miller, E. K. (2003). From rule to response: Neuronal processes in the premotor and prefrontal cortex. *Journal of Neurophysiology, 90*(3), 1790-1806.

Wessinger, C., VanMeter, J., Tian, B., Van Lare, J., Pekar, J., & Rauschecker, J. (2001). Hierarchical organization of the human auditory cortex revealed by functional magnetic resonance imaging. *Journal of Cognitive Neuroscience, 13*(1), 1-7.

Wilson, M. (2006). *Wandering significance. An essay in conceptual behavior.* Oxford: Clarendon Press.

Zerilli, J. (2017). Against the "system" module. *Philosophical Psychology, 30*(3), 235-250.