Integrative modeling and the role of neural constraints

Daniel A. Weiskopf

Abstract: Neuroscience constrains psychology, but stating these constraints with precision is not simple. Here I address the question of whether mechanistic analysis provides a useful way to integrate models of cognitive and neural structure. Recent evidence suggests that cognitive systems map onto overlapping, distributed networks of brain regions. These highly entangled networks often depart from stereotypical mechanistic behaviors. While this casts doubt on the prospects for classical mechanistic integration of psychology and neuroscience, I argue that it does not impugn a realistic interpretation of either type of model. Cognitive and neural models may depict different, but equally real, causal structures within the mind/brain.

1. The many-models problem

Constructing scientific models requires making choices about how to represent the structural, functional, and dynamical properties of whatever lies within the target domain. But without a way of modeling all of the world's aspects within a single comprehensive scheme, we are forced to find ways of coordinating and reconciling many incomplete models. The *many-models problem* arises when two or more substantially different models are applied to a single system in a context where their differences give rise to tensions about how to understand the real structure of the system itself. These contexts are ones in which models are interpreted in a minimally realistic way: they are intended to capture real entities, processes, and causal structures that are responsible for the observed phenomena. As Margaret Morrison (2000) notes, such an "incompatible models" scenario seems to pose a problem for scientific realism: unless the world can contain contradictory states, not all of these models can be strictly and literally true of the target system.

Here I consider one attempt to solve the many-models problem as it arises in cognitive neuroscience by using mechanistic analysis as a strategy of interfield integration. I cast a skeptical eye on this strategy by reviewing some evidence that psychological systems may not map onto neural systems in ways that are straightforwardly mechanistic, and I argue that such failures of integration do not show that these psychological systems are in any sense unreal.

2. Cognitive modeling strategies

Cognitive modeling aims to explain our psychological capacities for executing particular tasks, such as the ability to retain information in memory for short periods of time, or to categorize visually presented objects. Explaining these capacities requires describing the structure of the cognitive systems that underlie them, and tying the operations of those systems to the observed phenomena. These systems are recursively analyzed into interconnected subsystems, each of which is defined by its role in generating, transforming, and processing representations. A *cognitive model* is a depiction of how an ensemble of such systems interact to generate the target capacity or ability. Since capacities are identified in terms of their causal profiles, a cognitive model can explain the presence and exercise of a capacity only if the systems it depicts are capable of producing a matching profile.

These models are intended to capture part of the causal structure of the mind in terms of the coordinated operations of complex systems. This makes cognitive models a species of *componential causal model*, specifically one in which the elements of the model are characterized primarily in functional terms (Weiskopf, 2011). Insofar as these models have psychological systems, states, entities, and processes as their intended domain of reference, they

make no essential mention of the details of their material composition or the spatiotemporal organization of their components. An *accurate* cognitive model is one that depicts a set of real psychological structures that actually produce the target capacity in the subjects being studied. The accuracy of a cognitive model turns on whether the structure that it depicts is instantiated in the subjects being studied.

Several strategies can be used to determine whether a model is accurate. One involves applying model-fitting procedures to determine whether the model can capture the data that have been produced so far, and whether it generalizes to new data sets (see De Schutter, 2010 for examples). A second is to experimentally determine whether the constructs of the model are robust, where a robust construct is one that is detectable through a variety of independent epistemic channels, and manipulable through a variety of distinct interventions. A third way to assess the accuracy of models in psychology has gained prominence with the rise of cognitive neuroscience. This method involves determining how well these models can be integrated with other models of the same system. This is often expressed as the demand that psychological models should be *neurally plausible*.

Neural plausibility gained currency in the debate between classical and connectionist cognitive architectures, in which connectionist models were often taken to be more neurally plausible than classical ones on the grounds that the structure of networks is broadly similar to the anatomical and physiological organization of actual neural systems (Butler, 1994). On this reading, a cognitive model is neurally plausible to the extent that its structure and processing resemble what takes place in neural systems: units map onto neurons (or clusters thereof), connections onto axons and dendrites (or fiber tracts), activation onto action potentials and

graded potentials, and so on. This mapping abstracts away some details, but projects the broad structural characteristics of brain organization onto cognition.

As classical cognitivists objected, however, this is a needlessly strong requirement: while an account of neural plausibility requires imposing some constraints on how the elements of each model ought to be mapped onto one another, psychological structure does not need to be transparently recoverable from neural structure (McLaughlin & Warfield, 1994). But recent work on mechanistic modeling offers an alternative construal of neural plausibility.

3. Integration through mechanistic modeling

Mechanistic modeling can be regarded as a paradigm for understanding interlevel relations in complex systems. Its explanatory target is the means by which S carries out Φ. To take one much-discussed example, the discovery that the dentate gyrus of the hippocampus (the target structure S) was a locus of long-term potentiation (the function Φ) guided investigations into the molecular mechanisms of the granule cell synapses where it occurred, resulting in, among other things, the discovery of the crucial role that the influx of Ca^{2+} ions plays in regulating LTP (Craver, 2005).

Here a specific activity was found to take place in an anatomically and cytoarchitecturally circumscribed region, which was then decomposed to reveal its causally relevant parts, namely the particular synaptic structures and processes that produce LTP in dentate gyrus neurons. The resulting model puts together two spatiotemporal scales at which the same entity can be described. As a coherent whole, S carries out a certain function Φ, and at the same time the mechanism by which S Φ's is composed of the entities and processes that are

mereologically contained within the boundaries of the whole entity, and organized to causally support that function.

This suggests a way of cashing out neural plausibility: a cognitive model is plausible to the extent that it can be mechanistically integrated with a neural model. The idea that psychology should ultimately be supported by mechanistic integration with neuroscience lies behind the "3M Constraint" proposed by Kaplan and Craver (2011, p. 611): "In successful explanatory models in cognitive and systems neuroscience (a) the variables in the model correspond to components, activities, properties, and organizational features of the target mechanism that produces, maintains, or underlies the phenomenon, and (b) the (perhaps mathematical) dependencies posited among these variables in the model correspond to the (perhaps quantifiable) causal relations among the components of the target mechanism."

Mechanistic integration requires mapping cognitive systems, representations, and processes onto the elements and activities of neural mechanisms. Here the twin heuristics of decomposition and localization come into play. Functional decomposition involves the analysis of a system into its functionally relevant subparts; structural decomposition involves doing the same for its physical parts. In the case of interfield modeling, the job of functional analysis has already been carried out by the cognitive modeler. Localization requires finding a one-to-one assignment of elements of the cognitive model to distinct structural elements of the neural model, subject to the added constraint that these neural elements should be both spatially circumscribed and relatively "natural"—that is, not arbitrary or gerrymandered, from the point of view of our background theory of how the brain is organized. If localization is successful, cognitive functions will end up being assigned to distinct spatially and structurally well-defined

components of the brain. Everything that appears as a distinct element in the cognitive model will correspond to a distinct element of some neural mechanism.

It bears emphasis that localization of function is a significant constraint for mechanists (Silberstein & Chemero, 2014). The guiding image of mechanisms as machinelike structures strongly suggests that they are made of discrete parts each of which carries out a dedicated function. Mass-produced artifacts with this sort of design allow parts to be detached and swapped out without disrupting the organization of the whole. The less well-localized these functions become, and the less easily the parts can be separated from one another while retaining their own functions, the further the system drifts away from being mechanistic. This underlies both Simon's notion of near-decomposability and James Woodward's related constraint of functional modularity (Woodward, 2013).

Localization failures can arise in several ways. One possibility is that the posited function might have no corresponding structural element at all. A second is that functional assignments may overlap: if several functions are assigned to the same structural element, the specialization of functional parts that mechanism requires is violated. Localization of functions, then, implies that the neural elements that correspond to particular cognitive elements ought to be wholly or largely non-overlapping.

There are several reasons why localization of function significantly constrains interfield modeling. One is that ontologically distinct entities are expected to be *independently modifiable*. In a situation where mechanistic components overlap or interpenetrate, however, there may be a single local intervention that could affect two or more psychological systems simultaneously by targeting their region of overlap. If neural elements interpenetrate or share parts to any

significant degree, modifying one can potentially modify the other, and this casts doubt on the claim that the cognitive elements that map onto them are really distinct entities.

Another reason comes from the practice of *reverse inference*. A reverse inference is one that moves from the detection of a pattern of neural activity to the conclusion that a certain psychological state or process is taking place. Such inferences proceed on the assumption that psychological entities can be correctly discriminated by corresponding neural entities; that is, that neural entities are *selective*. As Russell Poldrack puts it, "If a region is activated by a large number of cognitive processes, then activation in that region provides relatively weak evidence of the engagement of the cognitive process; conversely, if the region is activated relatively selectively by the specific process of interest, then one can infer with substantial confidence that the process is engaged given activation in the region" (Poldrack, 2006, p. 2). The greater the degree of overlap between neural entities, the less distinguishable they will be, and the less powerful reverse inferences will become.

Poldrack has strengthened this proposal into a method for deciding on the legitimacy of elements of a cognitive model (Poldrack, 2010): "if it is not possible to distinguish two [psychological] concepts from one another (but it is possible to distinguish both from a different process), this suggests that the ontological distinction between those two concepts should be reconsidered" (p. 760). This method of distinguishing between psychological constructs involves comparing the overlap between the neural regions recruited in various tasks. If two tasks that purportedly employ different cognitive processes cannot be cleanly distinguished from each other in terms of neural activation, then the distinction between those processes is one that is potentially unreliable and on the table for revision.

As a "proof of concept" for this strategy, Lenartowicz et al. (2010) surveyed the imaging literature to see whether psychological processes associated with cognitive control could be neurally discriminated from one another. They searched the literature for a set of labels for different processes associated with cognitive control and retrieved from the BrainMap database the regions that manifested peaks of activity on tasks invoking those processes. They then constructed a pairwise comparison matrix showing the degree to which these patterns could be discriminated from one another. While most of the processes were fairly discriminable, task switching and response selection were poorly discriminated, as were task switching and response inhibition. Since the latter two processes are independently well discriminated from the remaining ones, task switching's weak pairwise discrimination may indicate that it is not a neurally plausible construct.

If cognitive functions are localized in separate, naturally circumscribed neural regions, they will be independently modifiable under some class of neural interventions, and will also support fairly strong reverse inferences. If, on the other hand, cognitive functions map onto interpenetrating, distributed neural elements, neither of these will necessarily hold. The latter scenario would also cast doubt on the possibility of a mechanistic integration of psychology and neuroscience, given the centrality of localization to this enterprise. I now turn to some evidence that this scenario ought to be taken seriously.

4. Network analysis and massive neural redeployment

Much current theoretical work in neuroscience makes use of network-analytic tools and methods to uncover brain structure at a variety of spatial scales, as well as the dominant modes

of large-scale neurophysiological organization (Sporns, 2011). A recurring theme in this research is that the primary unit for understanding brain function is whole networks, rather than individual localized areas. These networks consist of anatomically distinct regions that are linked by relatively persistent fiber pathways which support their tendencies to co-activate. Local structures contribute their specific processing capacity to these networks, but by themselves tend not to realize cognitively significant functions. Instead, these functions can only be assigned to the coordinated dynamical activity of large regions of cortex.

There are three facts about network structure that are significant here. The first is that the networks that are involved in cognitive processing have significant areas of anatomical overlap. The individual regions that play a role in one network may also play a role in many other networks, although there is no guarantee that the role that a region plays in one network is the same as the role it plays in another.

The second is that the same anatomical network can support many different dynamic functional configurations. The shifting activation dynamics of a network reflect changing task demands, stimulus characteristics, and changes to the brain's background processing conditions. This suggests that particular regions are multifunctional in two ways. Not only can they contribute processing to partially anatomically disjoint structural networks, they can also support different modes of processing within one and the same anatomical network. Studies of neural dynamics suggest that there are many such overlapping networks present, and that their activity is sensitive to ongoing cognitive tasks and context (Wig, Schlaggar, & Petersen, 2011).

The third is that what role a region plays, and how regions are functionally connected to one another, is often determined by nonlocal factors; that is, by factors that are extrinsic to the

region itself. Even if a region such as inferotemporal cortex is activated in two task contexts, what it is doing (object recognition vs. face recognition) in those contexts may differ based on the pattern of other regions that are coactivated in each one (Friston, 1997). Given the density of interconnections among brain areas, it would not be surprising if such extrinsic modulations of regional functions were commonplace.

This emphasis on the role of dynamic networks dovetails nicely with Michael Anderson's hypothesis of massive neural redeployment (Anderson, 2007a, 2007b, 2010, 2014). According to this view, cognitive functions depend on the coordinated activity of many different neural regions that are themselves highly multifunctional. Support from the thesis comes from several meta-analytic studies. In an early review of 35 PET studies by Dan Lloyd (2000), it was reported that separate cognitive tasks recruit on average 3.3 Brodmann areas, and each BA is implicated in 3.4 tasks. Another analysis of 135 tasks used in dozens of studies of attention, perception, imagery, and language suggests that these tasks involve an average of 5.97 different anatomical areas each. And these areas themselves participate in many other unrelated tasks: e.g., 93% of left lateralized regions participate in at least one other task of a different category (Cabeza & Nyberg, 2000).

Other reviews point in a similar direction. A review by Anderson and Pessoa (2011) examined the participation of 78 distinct anatomical regions across 1138 experimental tasks, which were each assigned to 11 possible BrainMap task domains. A diversity score was calculated for each region, representing the degree to which each area participated in different BrainMap tasks. The average diversity across all regions was .70, suggesting that they tended to be invoked in a large number of task domains. More interestingly, Anderson and colleagues derived a functional connectivity matrix for brain regions employed in 1127 experimental tasks

(Anderson & Penner-Wilger, 2013). Where regions are mutually active more often than would be predicted based on the probability of their individual activations, they are considered to be functionally connected. These coactivation patterns were used to generate network maps of brain regions. In the empirically derived networks, there is more overlap in nodes than in edges; i.e., the networks reuse the same anatomical components, but coordinate and deploy them in different ways. These graphs can be compared to the organization of "random" networks with the same numbers of edges and nodes to show that the functional networks have more node overlap and less edge overlap than would have been predicted by chance.

This organization is consistent with Sporns' claim that "the same set of network elements can participate in multiple cognitive functions by rapid reconfigurations of network links or functional connections" (2011, pp. 182-3). In all of these cases, the mapping from cognitive to neural organization is one in which many brain regions are linked with a single cognitive function, and a single region can participate in many functions. This is not just a many-many mapping, which implies that the realization base for cognitive systems will be "smeared out" and interdigitated, but it is also often overlapping. In an extreme case, two cognitive systems might share nearly all of a set of underlying brain regions; here, the regions would be ones in which dynamic reconfiguration of regional activity gives rise to different functions. One upshot of this form of organization is that the neural regions that participate in this assembly may have no identifiable cognitive function outside of their role in the ensemble.

While classical localization assumed that distinct cognitive systems would have disjoint physical realization bases, massive redeployment and network theory seem to demonstrate that different systems may have *entangled realizers*: shared physical structures spread out over a large region of cortex. This suggests that not only will there not be *distinct* mechanisms

corresponding to many of the systems depicted in otherwise well-supported cognitive models, but given that the relevant anatomical structures are multifunctional in a highly context-sensitive way, perhaps nothing much like mechanisms at all. And while it might be that these networks should count as mechanisms on a sufficiently liberal conception of what that involves, widespread entanglement still violates Poldrack's constraint that distinct cognitive structures should be realized in distinct neural structures. Given the centrality of the distinct localization constraint to cognitive neuroscience (Coltheart, 2001), it is worthwhile to explore what dropping it might entail.

5. Realistic model pluralism

Assuming that the foregoing claims about neural structure are correct, we seem to have a classic instance of the many-models problem, in which psychology and neuroscience can give rise to interestingly distinct models of the same system. While cognitive models depict one sort of causal structure for the mind/brain, network and massive redeployment models depict a different one, in which two or more ontologically distinct cognitive systems map onto a constellation of shared, overlapping brain networks. If we assume that no system can have more than one causal organization, we would be forced to the conclusion that one of these models must be false—perhaps that the neurally implausible cognitive model should be revised, or interpreted in a less than completely realistic fashion.

I contend that it is the assumption of a unique causal structure that should go instead. This depends on fleshing out a notion of what it means for a model to depict a causal structure. Here I will adopt some terminology from John Campbell (Campbell, 2006, 2008, 2010) and say

that the elements of a model (cognitive or neural) constitute *control variables* for the behavior of the system as a whole. To call an element a control variable is to analogize it to a dial, knob, lever, or switch on a control panel. Just as these particular physical control structures can, when attached to a well-designed system and appropriately manipulated, produce predictable outcomes on the part of the machine that they are coupled to, control variables within models stand for components that can be intervened on or manipulated to produce effects in a similar smoothly predictable fashion.

In more precise terms, Campbell considers something to be a control variable when: (1) there is a natural-seeming function from the variable to a set of possible outcomes; (2) changes in the variable can make a large difference in the possible outcome that is achieved; (3) these differences are largely specific to the outcome in question; and (4) there is a way of systematically manipulating or changing the variable. If all of these conditions are met for model components, then *prima facie* they depict genuine, causally active entities. If manipulating a component produces only unsystematic, negligible, or unspecific effects on a system's behavior, its claim to causal reality is weakened.

Much of the work of designing experiments in psychology, particularly where this work is guided by an existing model, can be interpreted as aiming at discovering control variables (Weiskopf, forthcoming). For example, components of cognitive models are defined in terms of their role in carrying out specific kinds of information processing tasks defined over materials having a particular format and content. Studies involving materials manipulations, then, are aimed at seeing how these components systematically change their behavior when exposed to a range of inputs. If their behavior is stable, predictable, and in line with what the model predicts about the component's role in the organization, this counts as reason to treat that component as a

control variable. The same goes for interference or dual-task experimental designs, which are intended to isolate particular components of a model by overloading them to see what effect their inhibition has on processing. In these paradigms, tasks that place high demands on a component are predicted to deactivate that component, which allows testing of the model's predictions of how the system will operate without it.

So a cognitive model can be said to be empirically well-validated when there is a range of experimental tasks that target each of its component elements, their processing structure, and their interconnections, and these tasks are sufficient to establish that these elements are control variables; that is, they can all be specifically and systematically manipulated to produce a range of well-defined effects. In this case, there is every reason to think that these elements are depicting causally real structures within the mind.

By the same token, models of neural structure and function may be treated as depicting proposals about the control variables for the dynamics of neurophysiology. When considering the behavior of a dynamic network, the nodes and edges provide potential local regions where we can intervene in the system via transient or longer-lasting procedures to see how its behavior changes at the collective level. Interventions such as lesions, stimulation by electrodes, or transient regional deactivation via transcranial magnetic stimulation lead to predictions about how the other nodes in the network will reconfigure themselves as a result. This has been modeled computationally using lesions in realistic network simulations, which indicate that localized damage can have both regional and more widespread effects on network function (Alstott, Breakspear, Hagmann, Cammoun, & Sporns, 2009).

However, the fact that there are two sets of control variables that can be used to model a system doesn't entail that they can be straightforwardly integrated. In particular, there may be no simple way of systematically manipulating a system's cognitive properties by designing interventions into its neural architecture. This is a characteristic shared by control variables in other domains (Campbell, 2008, pp. 437-442). Given the differences between cognitive and neural models, the control variables they provide may not align with each other. A single system that serves as a locus of cognitive interventions may not be uniquely, systematically targetable by a local neural intervention if cognitive systems generally have entangled realization bases. This is the sense in which these models depict *different* causal structures. However, if the causal structure depicted by each model is empirically validated, there are no grounds for preferring one to the other, or forcing a revision to one set of categories in favor of another.

Obviously these comments are speculative, in light of how little is yet known about the relationship between cognitive models and network-style neurophysiological models. Given the facts of model pluralism, however, we need not expect any neat relationship between the causal patterns picked out by these two types of models. Using the lens of cognitive modeling involves representing the system's functional components in one way, while using the lens of neurophysiology involves representing them in another. The essence of autonomy is that these lenses may simply produce different images.

6. Conclusions

This discussion, brief as it is, highlights some of the complications of interfield modeling. While many systems may turn out to be tractable using the techniques of mechanistic analysis, it

is less clear that psychological models can be integrated with neural models in this way. This isn't to deny that many aspects of brain function can be captured mechanistically, just that understanding the neural basis of cognitive functions may require a new set of analytical tools to deal with cases in which they are realized in a distributed, non-local, and entangled fashion. These cases should also signal a need for humility in applying constraints such as neural plausibility, especially insofar as our understanding of what is and isn't plausible is conditioned by thinking about this mapping in terms of assignment of cognitive functions to unique mechanistic structures. The mind-brain relationship might be considerably more opaque than that.

References

Alstott, J., Breakspear, M., Hagmann, P., Cammoun, L., & Sporns, O. (2009). Modeling the impact of lesions in the human brain. *PLoS Computational Biology*, *5*(6), 1-12.

Anderson, M. L. (2007a). Massive redeployment, exaptation, and the functional integration of cognitive operations. *Synthese*, *159*(3), 329–345.

Anderson, M. L. (2007b). The massive redeployment hypothesis and the functional topography of the brain. *Philosophical Psychology*, *20*(2), 143–174.

Anderson, M. L. (2010). Neural reuse: a fundamental organizational principle of the brain. *The Behavioral and Brain Sciences*, *33*(4), 245–66; discussion 266–313.

Anderson, M. L. (2014). *After Phrenology*. Cambridge, MA: MIT Press.

Anderson, M. L., & Penner-Wilger, M. (2013). Neural reuse in the evolution and development of the brain: evidence for developmental homology? *Developmental Psychobiology*, *55*(1), 42–51.

Butler, K. (1994). Neural constraints in cognitive science. *Minds and Machines*, *4*(2), 129–162.

Cabeza, R., & Nyberg, L. (2000). Imaging Cognition II: An Empirical Review of 275 PET and fMRI Studies. *Journal of Cognitive Neuroscience*, *12*(1), 1–47.

Campbell, J. (2006). An interventionist approach to causation in psychology. In A. Gopnik & L. E. Schulz (Eds.), *Causal learning: Psychology, philosophy, and computation* (pp. 58–66). Oxford, UK: Oxford University Press.

Campbell, J. (2008). Interventionism, control variables and causation in the qualitative world. *Philosophical Issues*, *18*(1), 426–445.

Campbell, J. (2010). Control Variables and Mental Causation. *Proceedings of the Aristotelian Society*, *110*, 15–30.

Cotlheart, M. (2001). Assumptions and methods in cognitive neuropsychology. In B. Rapp (Ed.), *Handbook of Cognitive Neuropsychology* (pp. 3-21). Philadelphia: Psychology Press.

Craver, C. F. (2005). Beyond reduction: mechanisms, multifield integration and the unity of neuroscience. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, *36*(2), 373–395.

De Schutter, E. (Ed.). (2010). *Computational Modeling Methods for Neuroscientists*. Cambridge, MA: MIT Press.

Friston, K. J. (1997). Imaging cognitive anatomy. *Trends in Cognitive Sciences*, *1*(1), 21–27.

Kaplan, D. M., & Craver, C. F. (2011). The Explanatory Force of Dynamical and Mathematical Models in Neuroscience: A Mechanistic Perspective. *Philosophy of Science*, *78*(Oct.), 601–627.

Lenartowicz, A., Kalar, D, Congdon, E., & Poldrack, R. A.. (2010). Towards an ontology of cognitive control. *Topics in Cognitive Science*, *2*(4), 678–692.

Lloyd, D. (2000). Terra Cognita: From Functional Neuroimaging to the Map of the Mind. *Brain and Mind*, *1*, 93–116.

McLaughlin, B. P., & Warfield, T. A. (1994). The allure of connectionism reexamined. *Synthese*, *101*(3), 365–400.

Morrison, M. (2000). *Unifying Scientific Theories*. Cambridge: Cambridge University Press.

Poldrack, R. A. (2006). Can cognitive processes be inferred from neuroimaging data? *Trends in Cognitive Sciences*, *10*(2), 59–63.

Poldrack, R. A. (2010). Mapping Mental Function to Brain Structure: How Can Cognitive Neuroimaging Succeed? *Perspectives on Psychological Science*, *5*(6), 753–761.

Silberstein, M., & Chemero, A. (2014). Constraints on Localization and Decomposition as Explanatory Strategies in the Biological Sciences, *80*(5), 958–970.

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

Weiskopf, D. A. (forthcoming). The explanatory autonomy of cognitive models. In D. M. Kaplan (Ed.), *Integrating Psychology and Neuroscience: Problems and Prospects*.

Weiskopf, D. A. (2011). Models and mechanisms in psychological explanation. *Synthese, 183*, 313–338.

Wig, G. S., Schlaggar, B. L., & Petersen, S. E. (2011). Concepts and principles in the analysis of brain networks. *Annals of the New York Academy of Sciences*, *1224*, 126–46.

Woodward, J. (2013). Mechanistic Explanation: Its Scope and Limits. *Aristotelian Society Supplementary Volume*, *87*(1), 39–65.