Carlos Zednik
(University of Osnabrück)

Systems, networks, and mechanistic explanations in neuroscience

Explanations in neuroscience are generally regarded as mechanistic explanations: they describe the particular "set of entities and activities organized such that they exhibit the phenomenon to be explained" (Craver, 2007, p. 5). Although this consensus is grounded in a series of publications that focus on molecular and cellular neuroscience (e.g. Craver, 2007; Machamer, Darden, & Craver, 2000), it has recently been extended to other branches of the discipline such as computational, cognitive, and behavioral neuroscience (e.g. Kaplan, 2011; Piccinini & Craver, 2011). In comparison, philosophical discussions have generally neglected the areas of systems and network neuroscience. By invoking novel imaging techniques and increasingly sophisticated computer models, systems and network neuroscientists study structural (i.e. anatomical) networks, functional networks of statistical interdependencies, and networks of causal interactions that exist at the level of individual neurons, neural populations, and cortical regions. Whereas some researchers in these sub-disciplines describe the individual nodes and connections of particular networks in the brain, others identify global principles of network organization, such as the degree of clustering or hierarchical modularity, or the presence and absence of "hub" nodes (Sporns, 2011). Yet other researchers develop mathematical models of a network's global dynamics to describe patterns of regular oscillation, stable activity over time, or temporal synchronization between parts of the network (Hemmen & Sejnowski, 2006).

To what extent are systems and network neuroscientists also engaged in the discovery and description of mechanisms? Although there is an intuitive sense in which brain networks are mechanisms, it is unclear whether these mechanisms can be discovered and described for the purposes of mechanistic explanation. As Bechtel & Richardson have argued,

> "While network models are not classical mechanistic models, there is still a clear sense in which they are mechanistic. The behavior of the system is a product of the

> activities occurring within it. All the components are simple mechanical units, and their interactions are all characterized in simple mechanical terms…There is clearly no case here for abandoning a mechanistic perspective. Nonetheless, these systems defy the approach to mechanism that we charted…because these systems are neither decomposable nor even minimally decomposable, and systematic functions cannot be localized." (Bechtel & Richardson, 1993, p. 228)

As it is traditionally conceived, mechanistic explanations are delivered by descriptions that adequately represent the individual component parts and operations of the mechanisms for the phenomena being explained, as well as their organization. Given the large-scale complexity of many brain networks, however, it is questionable whether such descriptions will ever be forthcoming. Indeed, even if the individual parts of a particular network can be identified, it seems difficult to determine each part's operation, and thereby difficult to reveal how that part contributes to the behavior of the network as a whole.

Bechtel & Richardson's assessment—echoed by many similar statements in the relevant literature—suggests two ways of construing the explanatory aspirations of systems and network neuroscientists. First, they may seek non-mechanistic explanations. Indeed, some mathematical models that describe a network's global dynamics have been construed as covering law explanations rather than mechanistic explanations (see e.g. Chemero & Silberstein, 2008). Second, systems and network neuroscientists might not seek explanations at all, as opposed to detailed descriptions. Insofar as covering law explanations are often accused of being "merely descriptive" rather than "genuinely explanatory" (see e.g. Kaplan & Craver, 2011), perhaps the aims of network neuroscientists are more modest than those of other branches of neuroscience? Indeed, many of the most well-funded research efforts such as the Human Connectome Project (Toga, Thompson, & van Horn, 2012), seem to aim primarily at the enumeration and description of brain networks, as opposed to showing how such networks might be linked to specific cognitive or behavioral phenomena.

Both construals are misleading. Bechtel & Richardson's assessment notwithstanding, systems and network neuroscience can in fact be construed as seeking mechanistic explanations. Consider the (largely completed) effort to describe the complete C. elegans nervous system—an effort analogous to the Human Connectome Project. Recently, Izquierdo & Beer (2013) have invoked graph-theoretical measures to derive, from a graphical representation of the C. *elegans* connectome, a "minimal network" of neurons that contribute to production of klinotaxis, a particular chemical gradient-following behavior. Thus, they invoke principled graph-theoretical means to identify which parts and properties of a particular network are in fact the components of a mechanism responsible for a target phenomenon. Although issues of scale will have to be resolved when this graph-theoretical methodology is extended to more complex organisms, Izquierdo & Beer provide a proof of concept that network neuroscience can and does seek mechanistic explanations at least in certain instances.

Even those areas of systems and network neuroscience that develop mathematical models of a network's global dynamics can be construed mechanistically. Here, it is important to look closely at the way in which the variables contained in these models are interpreted. Although some models specify variables that describe observable features of the phenomena being explained, others specify variables that describe organizational properties of the network from which these phenomena arise. For example, the dynamical neural field theory (Schöner & Dineva, 2007) accounts for infant perseverative reaching with a mathematical model in which the value of a single cooperativity parameter determines whether reaches are more likely to be driven by perceptual inputs or by memory of previous reaches. The value of this parameter, importantly, denotes the degree to which the neurons within a particular population are mutually excitatory or inhibitory. Therefore, although the model shies away from describing individual component parts and operations of the mechanism for perseverative reaching, it does describe a property of that mechanism's global organization.

Re-construing the explanatory aspirations of systems and network neuroscience in this way is not only crucial for attaining a better understanding of this increasingly important area of neuroscientific research, but is also likely to inform the philosophical conception of mechanistic explanation. As can be observed in Bechtel & Richardson's statement, philosophers typically emphasize the description and discovery of individual parts and operations. In contrast, network neuroscience is primarily concerned with the way certain kinds of mechanisms—large networks in the brain—are organized. Thus, by paying closer attention to this kind of research, philosophers of science are likely to attain an improved understanding of this crucial but regularly overlooked aspect of mechanistic explanation.

References:

Bechtel, W., & Richardson, R. C. (1993). Discovering complexity: Decomposition and localization as strategies in scientific research. Princeton, NJ: Princeton University Press.

Chemero, A., & Silberstein, M. (2008). After the Philosophy of Mind: Replacing Scholasticism with Science. Philosophy of Science, 75(1), 1–27. doi:10.1086/587820

Craver, C. F. (2007). Explaining the Brain. Oxford: Oxford University Press.

Hemmen, J. L., & Sejnowski, T. J. (2006). 23 Problems in Systems Neuroscience. Oxford: Oxford University Press.

Izquierdo, E. J., & Beer, R. D. (2013). Connecting a Connectome to Behavior: An Ensemble of Neuroanatomical Models of C. elegans Klinotaxis. PLoS Computational Biology, 9(2). doi: 10.1371/journal.pcbi.1002890

Kaplan, D. M. (2011). Explanation and description in computational neuroscience. Synthese, 183, 339–373. doi:10.1007/s11229-011-9970-0

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

Machamer, P., Darden, L., & Craver, C. F. (2000). Thinking about mechanisms. Philosophy of Science, 67(1), 1–25.

Piccinini, G., & Craver, C. F. (2011). Integrating psychology and neuroscience: functional analyses as mechanism sketches. Synthese, 183, 283–311. doi:10.1007/s11229-011-9898-4

Schöner, G., & Dineva, E. (2007). Dynamic instabilities as mechanisms for emergence. Developmental science, 10(1), 69–74. doi:10.1111/j.1467-7687.2007.00566.x

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

Toga, A. W., Thompson, P. L., & van Horn, J. D. (2012). Mapping the Human Connectome. Neurosurgery,
71, 1–5.