Constraints on Localization and Decomposition as Explanatory Strategies in the Biological Sciences

Abstract. Several articles have recently appeared arguing that there really are no viable alternatives to mechanistic explanation in the biological sciences (Kaplan and Craver 2011; Kaplan and Bechtel 2011). This claim is meant to hold both in principle and in practice. The basic claim is that any explanation of a particular feature of a biological system, including dynamical explanations, must ultimately be grounded in mechanistic explanation. There are several variations on this theme, some stronger and some weaker. In order to avoid equivocation and miscommunication, in section 1 we will argue that mechanistic explanation is defined by localization and decomposition. In section 2 we will argue that systems neuroscience contains explanations that violate both localization and decomposition on any non-trivial construal of these concepts. Therefore, in section 3 we conclude the mechanistic model of explanation either needs to stretch to now include explanations wherein localization or decomposition fail, or acknowledge that there are counter-examples to mechanistic explanation in the biological sciences. We will also consider consequences and possible replies on the part of the mechanist in section 3.

1. Introduction. While there are many different accounts of mechanistic explanation, the basic idea is that a phenomenon has been explained when the responsible realizing or underlying mechanism has been identified. In particular, the relevant parts of the mechanism and the operations they perform must be identified, i.e., those parts/operations that maintain, produce, or underlie the phenomena in question (Bechtel 2010; Craver 2007; Machamer, Darden and Craver 2000; Kaplan and Craver 2011; Kaplan and Bechtel 2011). Whatever the particular account of mechanistic explanation on offer, it is clear that mechanistic explanation is supposed to be fundamental in the biological sciences, period. What is less clear is exactly what this explanatory axiom entails. What follows is a list of claims pertaining to dynamical and mathematical explanations in the biological sciences that some mechanistic thinkers assert are entailed by the mechanistic model:

1) Dynamical and mathematical explanations in systems neuroscience must be grounded in or reduced to mechanistic explanations (via localization and decomposition) to be explanatory.

2) Dynamical mechanisms are not an alternative to mechanistic explanation but a complement.

3) When dynamical and mathematical models do not describe mechanisms by appropriately mapping elements of the latter onto the former, then they provide no real explanation.

4) At this juncture, dynamical and mathematical models of explanation in biology not sufficiently grounded in mechanisms have nothing to offer but “predictivism” by way of

explanatory force. That is, critics of mechanistic explanation do not have a viable alternative research strategy or alternative conception of explanation on offer (Kaplan and Bechtel 2011; Kaplan and Craver 2011).

The mechanists in question claim that certain defenders of dynamical and mathematical explanation in the biological sciences violate 1-3 and are therefore guilty of 4 (Kaplan and Craver 2011; Kaplan and Bechtel 2011). We first we need to get clear on exactly how the "dynamicist" is being portrayed. Kaplan and Craver go after the "strong dynamicist and functionalist", which they characterize as follows, "In particular, we oppose strong dynamicist and functionalist views according to which mathematical and computational models can explain a phenomenon without embracing commitments about the causal mechanisms that produce, underlie, or maintain it" (2011, 603). The strong dynamicist and functionalist holds that "mechanistic explanation is no longer an appropriate goal for cognitive and systems neuroscience" (Ibid). And finally, "If these dynamicists are right, such models yield explanations in the total absence of commitments regarding the causal mechanisms that produce the cognitive or system behavior we seek to explain" (Ibid, 604). According to Kaplan and Craver then, the strong dynamicist abandons the mechanistic model of explanation and has nothing coherent or cogent to replace it with.

We also reject strong dynamicism and functionalism so characterized. We will show however that 'either mechanistic explanation or dynamical predictivism' is a false dilemma. What we will claim is that systems biology and systems neuroscience contain robust dynamical and mathematical explanations of some phenomena in which the essential explanatory work is not be being done by localization and decomposition. More positively, the explanatory work in these models is being done by their graphical/network properties, geometric properties, or dynamical properties. We mean this claim to be true both in practice and in principle. Presumably then, what separates us from the mechanists is that they are committed to all such "higher level" explanations ultimately being discharged via localization and decomposition and we are not. However, we certainly do not think such explanations are incompatible or mutually exclusive, we have no problem calling them "complementary." Nonetheless, we will argue that graphical and dynamical properties for example are "non-decomposable" and non-localizable features of the causal and nomological structure of the "mechanisms" in question.

We want to end this section with a sociological note of caution. A great deal of the discussion in the literature strongly suggests that what we have before us is a thinly veiled iteration of the ancient philosophical debate between competing 'isms' regarding the essence of mind and the essence of explanation. Take the following, "It has not escaped our attention that 3M [mechanistic model of explanation], should it be found acceptable, has dire implications for functionalist theories of cognition that are not, ultimately, beholden to details about implementing mechanisms. We count this as significant progress in thinking about the explanatory aspirations of cognitive science" (Ibid, 612). So in one corner we have the functionalist/dynamist with their usual disregard/distaste for implementing mechanisms and in the other corner the mechanist, who insists on filling in all the boxes and the equations with the really truly fundamental

"causal structure." We think that it's time to transcend these beleaguered battle lines. That is, while we reject strong dynamicism and functionalism, and while we agree that dynamical and mechanistic explanations inevitably go hand-in-hand, we are open to the possibility that there are explanations in the biological sciences that are not best characterized in terms of localization and decomposition. To reject this possibility out of hand is as extreme as thinking that implementing mechanisms are irrelevant for explaining cognition and behavior.

When Kaplan and Craver say, "The mechanistic tradition should not be discarded lightly. After all, one of the grand achievements in the history of science has been to recognize that the diverse phenomena of our world yield to mechanistic explanation" (2011, 613), we agree. In fact, we don't think the mechanistic tradition should be discarded. What we do think is that the mechanistic tradition understood in terms of localization and decomposition is in principle not the only effective explanatory strategy in the life sciences.

2. Counter-Examples to Localization and Decomposition in Systems Neuroscience

2.1 *Defining Localization and Decomposition.* Localization and decomposition are universally regarded as the *sine qua non* of mechanistic explanation. Identifying the parts of a mechanism and their operations necessitates decomposing the mechanism. One can use different methods to decompose a mechanism functionally, into component operations, or structurally, into component parts (Bechtel and Richardson 2010). The ultimate goal is to line up the parts with the operations they perform, this is known as localization (Ibid). Proponents of mechanistic explanation like to emphasize the way it differs from the DN-model of explanation, which is based on laws. Mechanistic explanation is not about the derivation of phenomenon from initial conditions and dynamical laws, but rather explanation via localization and decomposition.

Mechanistic explanation is reductionist in the sense that explanation is in terms of the parts of the mechanism and the operations those parts perform. Parts and operations are at a lower level of organization than the mechanism as a whole. Bechtel says that the most conservative mechanistic account is one in which a mechanism is characterized as generating a phenomenon via a start-to-finish sequence of qualitatively characterized operations performed by identifiable component parts (2011, 534). However, Bechtel, Craver and others have recently emphasized how liberal mechanistic accounts have become. For example, Bechtel has stressed that the reductionist methodology of localization and decomposition must be "complemented" by contextualizing parts/operations both within a mechanism at a given level and between the mechanism and its environment at a higher level. The context in question includes spatial, temporal, causal, hierarchical and organizational.

We applaud and affirm the liberalization of mechanistic explanation. We assume, though, that these mechanists consider localization and decomposition as ultimately essential to mechanistic explanation. That said, we wonder what they would count as counter-examples in principle. Fortunately, Bechtel and Richardson (2010) give us some

clues. They emphasize that localization and decomposition are "heuristic" strategies that sometimes fail when a system fails to be decomposable or nearly decomposable (Ibid, 13). According to them, there are two kinds of failures of decomposability or localizability: 1) when there are no component parts or operations that can be distinguished (such as a connectionist network), in which case one can only talk about organizational features—the best one can hope for here is functional decomposition, and 2) when there are component parts and operations but their individual behaviors systematically and continuously affect one another in a non-linear fashion. In this case mechanisms are not sequential but have a cyclic organization rife with oscillations, feedback loops, or recurrent connections between components. In these instances there is a high-degree of interactivity among the components and the system is non-decomposable and therefore localization will fail (Ibid, 24). In addition, if the non-linearity affecting component operations also affects the behavior of the system as a whole, such that the component properties/states are dependent on a total state-independent characterization of the system (i.e., one sufficient to determine the state and the dynamics of the system as a whole), then the behavior of the system can be called "emergent" (Ibid, 25). They emphasize that when the feedback is system wide such that almost all "The operations of component parts in the system will depend on the actual behavior and the capacities of other its components" (Ibid, 24), the following obtains. First, the behavior of the component parts considered within the system as a whole are not predicable in principle from their behavior in isolation. Second, the behavior of the system as a whole cannot be predicted even in principle from the separable Hamiltonians of the component parts (Ibid).

We affirm all this and indeed others have stressed these points in illustrating the *limits* of localization and decomposition (Chemero and Silberstein 2008; Stepp, Chemero, and Turvey 2011). However, what puzzles us is that Bechtel and Richardson go on to say that, "When these conditions are met, the systemic behavior is reasonably counted as emergent, even though it is fully explicable mechanistically" (Ibid, 24). Here Bechtel and Richardson seem to be saying that even though such "emergent" behavior is not amenable to decomposition or localization, it is nonetheless mechanistically explicable. But, in exactly what sense are such systems *mechanistically* explicable? We shall return to this in section 3, after we consider explanations in systems neuroscience.

2.2 *Explanation in Systems Neuroscience.* Systems neuroscience is a rapidly growing area devoted to figuring out how the brain engages in the coordination and integration of distributed processes at the various length and time scales necessary for cognition and action. The assumption is that most of this coordination represents patterns of spontaneous, self-organizing, macroscopic spatiotemporal patterns which resemble the on-the-fly functional networks recruited during activity. This coordination often occurs at extremely fast time scales with short durations and rapid changes. There is a wide repertoire of models used to account for these self-orgainzing macroscopic patterns, such as oscillations, synchronization, metastability, and nonlinear dynamical coupling. Many explanatory models such as synergetics and neural dynamics combine several of these features, e.g., phase-locking among oscillations of different frequencies (Sporns 2011).

Despite the differences among these models, there are some important generalizations to be had. First, dynamic coordination is often highly distributed and non-local. Second, population coding, cooperative, or collective effects prevail. Third, time and timing is essential in a number of ways. Fourth, these processes exhibit both robustness and plasticity. Fifth, these processes are highly context and task sensitive. Regarding the third point, there is a growing consensus that such integrated processes are best viewed not as vectors of activity or neural signals, but as dynamically evolving graphs. The evidence suggests that standard neural codes such as rate codes and firing frequencies are insufficient to explain the rapid and rapidly transitioning coordination. Rather, the explanation must involve "temporal codes" or "temporal binding" such as spike timing-dependent plasticity wherein neural populations are bound by the simultaneity of firing and precise timing is essential. In these cases neurons are bound into a group or functional network as a function of synchronization in time. The key explanatory features of such models then involves various time-varying properties such as: the exact timing of a spike, the ordering or sequencing of processing events, the rich moment-to-moment context of real world activity and immediate stimulus environment, an individual's *history* such as that related to network activation and learning, etc. All of the above can be modeled as attractor states that constrain and bias the recruitment of brain networks during active tasks and behavior (Von der Malsberg et. al, 2010).

There is now a branch of systems neuroscience devoted to the application of network theory to the brain. The formal tools of network theory are graph theory and dynamical system theory, the latter to represent network dynamics—temporally evolving dynamical processes unfolding in various kinds of networks. While these techniques can be applied at any scale of brain activity, here we will be concerned with large-scale brain networks. These relatively new to neuroscience explanatory tools (i.e., simulations) are enabled by large data sets and increased computational power. The brain is modeled as a complex system: networks of (often non-linear) interacting components such as neurons, neural assemblies and brain regions. In these models, rather than viewing the neurons, cell groups or brain regions as the basic unit of explanation, it is brain multiscale networks and their large-scale, distributed and non-local connections or interactions that are the basic unit of explanation (Sporns 2011). The study of this integrative brain function and connectivity is primarily based in topological features (network architecture) of the network that are insensitive to, and multiply realizable with respect to, lower level neurochemical and wiring details. More specifically, a graph is a mathematical representation of some actual (in this case) biological many-bodied system. The nodes in these models represent neurons, cell populations, brain regions, etc., and the edges represent connections between the nodes. The edges can represent structural features such as synaptic pathways and other wiring diagram type features or they can represent more functional topological features such as graphical distance (as opposed to spatial distance).

Here we focus on the latter wherein the interest is in mapping the *interactions* (edges) between the local neighborhood networks, i.e., global topological features—the architecture of the brain as a whole. While there are local networks within networks, it is the global connection between these that is of greatest concern in systems neuroscience. Graph theory is replete with a zoo of different kinds of network topologies, but one of perhaps greatest interest to systems neuroscience are small-world networks as various

regions of the brain and the brain as a whole are known to instantiate such a network. The key topological properties of small-world networks are: 1) a much higher clustering coefficient relative to random networks with equal numbers of nodes and edges and 2) short (topological) path length. That is, small-world networks exhibit a high degree of *topological* modularity (not to be confused with anatomical or cognitive modularity) and non-local or long-range connectivity. Keep in mind that there are many different types of small-world networks with unique properties, some with more or less *topological* modularity, higher and lower degrees (as measured by the adjacency or connection matrix), etc. (Sporns 2011; Von der Malsberg et. al 2010).

The explanatory point is that such graphical simulations allow us to *derive, predict* and *discover* a number of important things such as mappings between structural and functional features of the brain, cognitive capacities, organizational features such as degeneracy, robustness and plasticity, structural or wiring diagram features, various pathologies such as schizophrenia, autism and other "connectivity disorders" when small-world networks are disrupted, and other essential kinds of brain coordination such as neural synchronization, etc. In each case, the evidence is that the mapping between structural and topological features is at least many-one. Very different neurochemical mechanisms and wiring diagrams can instantiate the same networks and thus perform the same cognitive functions. Indeed, it is primarily the *topologica*l features of various types of small-world networks that explain essential organizational features of brains, as opposed to *lower level, local* purely *structural* features. Structural and topological processes occur at radically different and hard (if not impossible) to relate time-scales. The behavior and distribution of various nodes such as local networks are determined by their non-local or global connections. As Sporns puts it, "Heterogeneous, multiscale patterns of structural connectivity [small-world networks] shape the functional interactions of neural units, the spreading of activation and the appearance of synchrony and coherence" (2011, 259).

Thanks to its generality and formal power, network neuroscience has also discovered various *predictive power laws* and *scale-free invariances,* i.e., symmetry principles at work in the brain. For example, the probability of finding a node with a degree twice as large as an arbitrary number decreases by a constant factor over the entire distribution. The explanatory power of small-world networks derives from their organizational properties, and not from the independent properties of the entities that are in small-world networks.

3. Consequences. Surprisingly, Bechtel and Richardson themselves use small-world networks as an example to illustrate that "mechanisms" of this sort require an addition to the mechanistic armament, namely, "dynamic mechanistic explanation" (Bechtel and Richardson, 2011, 16). Dynamical mechanistic explanation utilizes the tools of dynamical systems theory such as differential equations, network theory, etc., to engage in the computer simulation of complex mechanisms wherein the differential equations in question cannot be solved analytically. They claim of course that such "dynamical" explanations should nonetheless be squarely viewed as mechanistic explanation because:

> Reliance on simulations that use equations to understand the behavior of mechanisms may appear to depart from the mechanistic perspective and embrace

> something very much like the DN account of explanation. A simulation involves deriving values for variables at subsequent times from the equations and values at an initial time. However, simulations are crucially different from DN explanations. First, the equations are advanced not as general laws but as descriptions of the operations of specific parts of a mechanism. Second, the purpose of a computational simulation (like mental simulation in the basic mechanistic account) is not to derive the phenomenon being explained but to determine whether the proposed mechanism would exhibit the phenomenon. Finally, an important part of evaluating the adequacy of a computational model is that the parts and operations it describes are those that can be discovered through traditional techniques for decomposing mechanisms (Bechtel, 2011, 553).

There are several things that need to be said here. First, we agree that dynamical and network-type explanations are not D-N explanation and therefore cannot be guilty of "predictivism." Secondly, we agree that such explanations are nonetheless about *predicting* whether certain *causal structures* will have certain cognitive, functional or other features. Certainly, the fact that these simulations or dynamical/graphical systems predict or allow us to derive certain features does not make them explanatory. What does make them explanatory? These simulations show why certain *causal* and *nomological* structures will exhibit said features *in virtue of* their dynamical and graphical properties. Bechtel and company will balk at the word 'nomological', because the equations are not "advanced as general laws." When defending law-like explanations and the existence of laws in the special sciences, it is customary to point out that even the laws of physics do not always meet the ideals of the D-N model. That is, physical laws are often not spatiotemporally universal or free of exceptions, *ceteris paribus* clauses, idealizations and approximations. We are happpy however to forgo the word law in favor of Bechtel's phrase "organizational principles." For example, in network-based explanations the organizing principles include the aforementioned "power laws", involving self-similarity, scale-invariance and fractal patterning in space and time. Thirdly, while it may be true that one aspect of evaluating the adequacy of a computational model is that the parts and operations it describes are discovered through traditional techniques of decomposition, it should be clear that the brain networks being described here are non-decomposable and non-localizable. There is a degree of functional decomposition for these networks but not structural decomposition. That is, localization is simply beside the point.

There is no question that graphical and dynamical simulations do describe mechanisms, but they are not merely abstract descriptions of structural mechanisms. The key question here is what's really doing the explanatory work and the answer in this case is not in the structural or lower level mechanistic details. The simulations are not merely idealizations and approximations of such lower level structural interactions. Kaplan and Craver would claim that these models are mechanistic because they meet the "3M" criterion.

> 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 (2011, 611).

If what Kaplan and Craver mean to assert here is that any explanation proffered by a mathematical model of a mechanism is only truly explanatory if and only if said explanation can be reduced to or simply mapped onto the lower level structural features of the mechanism, then such mathematical models fail to be explanatory. Again, these graphical and dynamical models are non-decomposable and non-localizable. Otherwise, networks-based explanation easily meet the 3M criteria.

The key question is whether brains have the topological architectures they do in virtue of their structural mechanisms, or vice-versa? Or put another way, *in virtue of what* do graph theoretic models explain? As Bechtel himself admits, in such non-decomposable complex systems, the global topological features act as order-parameters (collective variables) that greatly constrain the behavior of the structural elements. As Sporns puts it, "a reentrant system operates less as a hierarchy and more as a heterarchy, where super- and subordinate levels are indistinct, most interactions are circular, and control is decentralized" (2011, 193). The dynamical interactions here are recurrent, recursive and reentrant. So there is no sense in which the arrow of explanation or determination is in principle exclusively from the 'lower level' structural to the 'higher' level graphical-dynamical. There is no structural, reductive or "downward-looking" explanation for the essential graphical properties of brain networks. Simply put, such *global* organizational principles or features of complex systems are not explicable in principle via localization and decomposition.

This is true for many reasons. The aforementioned many-one relationship between the structural and graphical features illustrates that specific structural features are neither necessary nor sufficient for determining global topological features. That is, topological features such as the properties of small-world networks exhibit a kind of "universality" with respect to lower level structural details. This is why in complex systems research part of the goal is to discover power laws and other scale-invariant relations. These laws allow us to predict and explain the behavior and future time evolution of the global state of the system regardless of its structural implementation. It turns out the reason power laws are predictive and unifying is that they show *why* the macroscopic dynamics and topological features obtain across diverse lower level structural details. And the *why* has nothing to do with similar structural details of the disparate systems.

A very brief and informal characterization of universality might be helpful here. There are many cases of universality in physics at diverse scales, but the general idea is that a number of microphysically heterogeneous systems, sometimes even obeying different fundamental equations of motion, end up exhibiting the same phenomenological behavior. When this happens we say such systems share the same critical exponents and thus all belong to the same universality class. The explanandum of universality is the uniformity and convergence of large-scale behavior across many very diverse instances.

That is, universality is a feature of classes of systems, not a specific system. The Renormalization Group analysis (RG) explains why specific physical systems divide into distinct universality classes in terms of the geometry or topology of the state space of systems, i.e., the so-called fixed points of the renormalization flow. Hamiltonians describing heterogeneous physical systems fall into the basin of attraction of the same renormalization group fixed point. The space of Hamiltonians contains numerous fixed points, each of which is describing different universality classes with different critical exponents and scaling functions. The microphysically diverse systems in the same universality class will exhibit a continuous phase transition, near which, their analogous macroscopic quantities will obey power laws possessing exactly the same numerical values of the critical exponents. The quantitative behavior near phase transitions exhibits this universality wherein the values of the exponents are identical.

What is interesting here is that techniques such as RG methods from statistical mechanics are being successfully applied to complex biological systems that don't have uniform parts. The occurrence of scale-invariance and hence self-similarity is the deeper reason why microphysically and mechanistically diverse systems can exhibit very similar or even identical macroscopic behavior. Thus, there is a direct route from power law behavior, scale-invariance and self-similarity to explaining why universality is true even in complex biological systems. Global topological features cannot be predicted from or derived *ab initio* from the structural features, because these are *qualitatively* different *types* of properties.

We take no position over whether these are genuine laws: we agree with Woodward (2003) that there is no need to determine whether something is a genuine law or a mere invariance to determine whether it can be used in explanation. The manner of explanation involved here is distinctly nomological. The laws found in systems neuroscience have more in common with laws found in physics than most special science laws. This is not surprising since the formal methods involved are mostly imported from physics. In fact, when it comes to the traditional virtues one expects of laws (e.g., quantifiability, universality, predictive power, satisfaction of counterfactual conditionals, explanatory power, simplicity, unification, etc.), the laws in systems neuroscience are no worse off than most laws in physics.

Explanations in systems neuroscience are highly pluralistic involving aspects of mechanistic, dynamical, various causal and statistical-causal explanations. Many *explanatory techniques* are used in this endeavor including a host of causal and statistical modeling techniques and variety of formal/statistical measures of complexity. There are various hybrids of these explanatory patterns as well. Therefore systems neuroscience embraces *explanatory and causal pluralism* as a matter of pragmatic explanatory practice. However, the norms of such systems neuroscience explanations decidedly transcend those of localization.

Following Woodward (2003), many mechanists such as Kaplan and Craver (2011) have adopted an interventionist account of mechanistic explanation in which a mechanistic explanation is only explanatory if it allows us to manipulate various "knobs and levers" of the mechanism thereby providing us with some control over the manifestation of the

phenomenon. Said control should allow us to "predict" how the system will behave if certain parts are broken, knocked-out, altered, etc. Kaplan and Craver allege that one of the things that separates dynamical explanations from real (causal) explanations, is that the former do not allow for intervention, manipulation or control. However, explanations in systems neuroscience are consistent with manipulationist or interventionist theories of explanation in general. Indeed, not just structural decompositions, but also dynamical and graphical explanations, can be and often are interventionist explanations. Mechanistic accounts of explanation that focus on localization and decomposition have no monopoly on interventionist explanation. There is nothing that says the knobs being tweaked must be structural components, they can also be global nomological features such as order-parameters or laws.

The kinds of complex biological systems under discussion here present a problem for any simplistic interventionist mechanistic model however. For example, often knock-out type experiments reveal that because of various types of plasticity, robustness/degeneracy and autonomy in complex biological systems, turning specific structural elements on or off, such as genes, has no discernable or predictable effect. In other words, we learn that such systems are non-decomposable and thus not amenable to localization. Needless to say, global organizational features such as plasticity, robustness, degeneracy and autonomy are not explicable via localization either. Therefore, very often the type of efficacious and informative manipulations one performs on such systems involves not structural components but global features such as order-parameters.

4. Conclusion. We have been arguing that the kinds of explanation common in systems neuroscience do not involve decomposition and localization. This would seem to make them non-mechanistic. It makes no difference us whether the mechanists want to stretch mechanistic explanation to include explanations wherein localization or decomposition fail, or whether they want to acknowledge that there are counter-examples to mechanistic explanation in systems neuroscience. We do think however that these are the only options remaining to the mechanist.

We have seen that: 1) there are mathematical explanations in systems neuroscience that are not grounded in localization and decomposition in principle, 2) mathematical explanations in systems neuroscience are complementary to explanations via localization and decomposition but not reducible to them, 3) while one can sometimes map structural elements onto mathematical explanations in systems neuroscience, the mapping is at least many-one and does not allow for structural decomposition or localization and 4) systems neuroscience really does provide an explanatory alternative to localization and decomposition that greatly transcends mere "predictivism."

References

Bechtel, W. (2009). Explanation: Mechanism, Modularity and Situated Cognition, in *The Cambridge Handbook of Situated Cognition.* P. Robbins and M. Aydede (eds.). Cambridge University Press

Bechtel, W. (2010). "Dynamic Mechanistic Explanation: Computational Modeling of Circadian Rhythms as an Exemplar for Cognitive Science", in *Studies in History and Philosophy of Science*. A 41:321-33.

Bechtel, W. (2011). "mechanism and Biological Explanation", in *Philosophy of science*. Volume 78:4. 533-558.

Bechtel, W. Richardson, R.C. (2010). *Discovering Complexity: Decomposition and Localization as Strategies in Scientific Research*. Second edition. Cambridge, MA: MIT Press/Bradford Books.

Chemero, A. Silberstein, M. (2008). "After the Philosophy of Mind: Replacing Scholasticism with Science" in *Philosophy of Science*. Volume 75, No. 1: 1-27.

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

Craver, C. Bechtel, W. (2007). "Top-Down Causation without Top-Down Causes", in *Biology and Philosophy* 22:547-63.

Kaplan, D. Bechtel, W. (2011). "Dynamical Models: An Alternative or Complement to Mechanistic Explanations?", in *Topics in Cognitive Science* 3. 438–444.

Kaplan, D. Craver, C. (2011). "The Explanatory Force of Dynamical and Mathematical Models in Neuroscience: A Mechanistic Perspective", in *Philosophy of science*. Volume 78:4. 601-28.

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

Stepp, N., A. Chemero, and M. Turvey. (2011). Philosophy for the rest of cognitive science. *Topics in Cognitive Science*, *3,* 425-437.

Von der Malsburg, C. Philips, W. Singer, W. (2010). *Dynamic Coordination and the Brain: from Neurons to Mind.* Cambridge: MIT Press.

Woodward, J. (2003). *Making Things Happen.* New York: Oxford University Press.