Chapter 9
Generalizing Mechanistic Explanations Using Graph-Theoretic Representations

William Bechtel

Abstract Mechanistic explanations appeal to the parts, operations, and organizations of mechanisms to explain the phenomena for which they are responsible. Scientists have developed accounts of myriads of mechanisms thought to be operative in biology, each involving distinctive parts and operations organized in idiosyncratic ways. The focus on specific mechanisms (e.g., those found in a particular cell type in a given model organism) appears opposed to the idea that explanations ought to be generalizable to new instances. Some generalizability can arise from the fact that many biological mechanisms inherit their parts and operations from mechanisms found in ancestral species and one can often identify commonalities in these parts and the operations they perform. But organization seems to be idiosyncratic to specific mechanisms, thwarting attempts to develop generalizations about how mechanisms organized in a specific way will behave. For example, as a result of different genes being expressed, the organizational pattern of interactions between proteins varies in different tissues or in the same tissue in different strains of a species. This poses an even greater problem when it is recognized that many biological mechanisms exhibit non-sequential organization of non-linear operations, making it difficult to use mental simulation to determine the behavior of the mechanism. Instead researchers resort to computational models, resulting in dynamic mechanistic explanations that integrate mathematical modeling with empirically ascertained details of parts and operations. These models, though, appear to be even more idiosyncratic, revealing only the behavior of the specific organization employed in a specific mechanism.

In recent years, however powerful tools have been developed for abstracting from the details of individual networks, providing a basis for informative generalizations about how networks employing the same abstract design will behave. These involve developing graph-theoretic representations of mechanisms and analyzing the properties of classes of graphs. Watts and Strogatz, Barabási, and Sporns have shown that large networks (e.g., neural circuits or gene networks) often exhibit

W. Bechtel (✉)
Department of Philosophy, Center for Circadian Biology, and Interdisciplinary Program in Cognitive Science, University of California, 9500 Gilman Drive, La Jolla, San Diego, CA, 92093-0119 USA
e-mail: bill@mechanism.ucsd.edu

P.-A. Braillard, C. Malaterre (eds.), *Explanation in Biology*, History, Philosophy and Theory of the Life Sciences 11, DOI 10.1007/978-94-017-9822-8_9

a scale-free, small-world organization capable of efficient, flexible coordination of operations and appeal to these properties to explain behaviors of specific mechanisms. Alon and Tyson, focusing on sub-networks with just two to four nodes, have identified different *motifs* (distinctive micro-architectures such as feedforward loops and double negative feedback loops) that are specialized for particular types of processing. These tools offer abstract organizational principles to which researchers appeal in their efforts to explain the behaviors generated by the mechanisms in which they are implemented. In this paper I show how these projects provide a basis for developing generalizable accounts of complex mechanisms and their dynamical behavior.

Keywords Dynamic mechanistic explanations • Generalizing explanations • Graph-theory formalizations • Motifs • Scale-free small worlds

1 Introduction: Formalizing Mechanisms in Graphs

In a broad range of biological disciplines the explanations researchers have pursued over the past two centuries have taken the form of characterizing the mechanisms responsible for various phenomena and demonstrating how they are able to generate the phenomenon. For example, metabolic research has described mechanisms for phenomena such as glycolysis or fatty acid metabolism, molecular biology has proposed mechanisms for cell division and protein synthesis, and neuroscience has investigated mechanisms for generating action potentials in neurons and for spatial navigation by animals. Although during much of the past two centuries philosophers had little to say about such explanations, over the last two decades several philosophers have advanced accounts of what mechanistic explanations are, how they are discovered and revised, and how they are evaluated (Bechtel and Richardson 1993, 2010; Bechtel 2011; Craver 2007; Glennan 1996, 2002; Machamer et al. 2000).[1]

A fundamental step in developing a mechanistic account is to characterize the phenomenon to be explained and to identify a mechanism taken to be responsible for its occurrence in the situations in which it occurs. This is often a non-trivial endeavor requiring experimental manipulation and quantitative analysis of the results. Research then seeks to decompose this mechanism into its components and operations. Biologists have developed powerful tools and inference strategies for decomposing mechanisms into their component parts and determining the operations for which these parts are responsible. But explanation also requires recomposing the mechanism to show that, in a given situation, it is able to produce the phenomenon (Bechtel 2011; Bechtel and Abrahamsen 2009). Recomposing the

[1]See Théry (2015, this volume) for a critical assessment of mechanistic explanations. See also Baetu (2015, this volume) and Brigandt (2013) about how mathematical models and mechanistic explanations may mutually support each other.

mechanism requires determining how the mechanism is organized and demonstrating that when so organized the operations identified are capable of jointly producing the phenomenon.

Researchers often approach the task of recomposition by assuming that the operations are organized sequentially. In such cases researchers can simulate the behavior of the mechanism mentally by representing the effects of each operation in the mechanism successively. In many cases, however, researchers come to recognize that sequential organization is insufficient—multiple operations may be performed simultaneously and individual components may have effects on other components responsible for other operations. The result is that individual components function differently depending on the context within the mechanism or beyond it. As various operations modulate the execution of other operations, the behavior of the whole mechanism may begin to produce a complex dynamic pattern such as oscillatory behavior or synchronization of independent oscillators. Mental simulation, even when assisted by a diagrammatic representation of the mechanism, commonly fails to provide accurate accounts of how such mechanisms will behave. In these circumstances it often proves highly useful to represent the mechanism in a computational model and to run computer simulations to determine how the mechanism will behave. I refer to explanations that rely on computational models to understand the dynamic behavior of a mechanism as *dynamic mechanistic explanations* (Bechtel and Abrahamsen 2010, 2011).[2] In the case of complicated mechanisms, these models may involve dozens or hundreds of differential equations. If appropriate parameter values are employed, these equations may together accurately simulate the behavior of the target mechanism. But often this accuracy comes at the expense of reducing the intelligibility of the explanation.[3]

A further problem with mechanistic or dynamic mechanistic explanations of mechanisms is that they are often developed for specific instances of the mechanism that have been investigated. Typically in biology mechanistic research is pursued on specific model systems or model organisms, which may not be the target of interest. A model organism may be chosen because of what is already known about the model organism, tools that have been developed for intervening on it, and the hope that the mechanisms in it will be simpler (although not simple) and so more tractable than

[2]While recognizing that one might integrate them, Woodward (2013) emphasizes the differences between more prototypical mechanistic approaches and those that adopt a more holistic perspective in, for example, explaining the robustness of a biological system or the dynamical behavior of a complex network. Woodward is right to emphasize the difference in research strategies and assumptions of the two approaches. More traditional mechanistic approaches assume modularity of the mechanism and dependence on the fine-tuned details of the system. In developing dynamical models of integrated systems researchers typically abstract away from these details. My reason for emphasizing their integration in dynamic mechanistic explanations is that many contemporary biological explanations draw on both details about the components of the mechanism and the dynamics of the integrated system, attempting to integrate the approaches that on their own make different assumptions and pursue different research strategies.

[3]See Issad and Malaterre (2015, this volume) for a discussion of the explanatory force of dynamic mechanistic explanations.

those in the target. This raises the question of whether the mechanistic or dynamic mechanistic account developed on the basis of research on the model system will generalize to the target system. Typically there will be differences between the model system and the target system, and researchers must assess whether these differences matter. Researchers working with computational models involving non-linearities recognize that often small changes lead to unpredicted major changes in the behavior of the model. This points to the challenge in generalizing from the model to the target.

To support generalizing results from the model to the target, researchers often appeal to the conservation of the mechanism through phylogeny—the mechanisms in contemporary are treated as variants that evolved through descent with modification from a common ancestor. When the model organism has evolved fewer additional components (as a result, for example, of gene duplications), it is expected to provide a less complicated version of the mechanism of interest in the target organism. The fact that the model is less complicated, however, points to differences between the model and the target, making it important to independently test the applicability of the explanation in the target system. The important issues in generalizing from model organisms are increasingly the focus of philosophical inquiry (Ankeny 2001; Ankeny and Leonelli 2011; Leonelli et al. 2014; Meunier 2012).

Another mode of generalization, which will be my focus here, is to abstract from the specific details of the mechanism to identify the pattern of organization (Levy and Bechtel 2013; Overton 2011). This has proven particularly valuable when the non-sequential organization of a mechanism results in dynamical behavior and researchers can explain that dynamical pattern as an expected consequence of a particular mode of organization. As I will demonstrate, computational modeling of the abstract mode of organization can contribute to the explanation of why mechanisms implementing the form of organization exhibit the behavior they do. But the key step is abstraction—leaving out details about the component parts and operations so as to focus only on the organization and determine its implications. Generalization then applies the information acquired through analyzing the abstract pattern of organization to different mechanisms that implement that pattern.

A particularly powerful means of identifying the abstract organization of a mechanism is to formalize the organization of the mechanism in graph-theoretical terms.[4] A graph consists of a collection of nodes connected by edges. The nodes represent entities, which in the case of a mechanism will include the component parts but may also include any other entities with which the mechanism or its parts interact. By treating different types of entities as nodes, a graph abstracts from the

[4]Jones (2014) discusses graph representations of the topology of a system that are not intended to be mechanistic but to identify such things as vulnerabilities of a system to disruption. While the diagrams he discusses and those discussed below often abstract from the details of the mechanism and thus provide a different type of understanding than do accounts of the parts and operations of a mechanism, they nonetheless play an important role in understanding how, as a result of its organization, a mechanism behaves either normally or when disrupted.

specific nature of the parts. Edges represent relations between nodes, which in many cases will be causal, although they may reflect correlations or even just the presence of physical connections. When the edges are causal, they are directional and in the case of a mechanism reflect the fact that an operation performed by one part or entity has effects on another. As in the case of nodes, edges abstract from the specific character of the operation.

As a basis for the discussion in subsequent sections, I present two examples in which, in the pursuit of research on a mechanism, researchers formalized the organization graph-theoretically. My goal at this stage is not to discuss the ways in which such graphs are analyzed, but simply to show contexts in which they are introduced. Working from slices examined with the confocal electron microscope, White et al. (1986) mapped the entire nervous system of the nematode worm *Caenorhabditis elegans*. The researchers identified 302 neurons and over 7,000 synapses. In the course of presenting their analysis, they employed many figures, including microscopic images, showing individual neurons and their projections in context. The last set of figures, however, abstracts from this detail and offers six circuit diagrams of the nervous system, of which the one relating amphids involved in chemoreception is shown on the top in Fig. 9.1. This circuit diagram does maintain some detail of the actual nervous system (chemical synapses are distinguished by arrows whereas gap junctions are represented by bar-ended lines, sensory neurons are shown as triangles and interneurons by hexagons, and the prominence of synapses is indicated by the number of cross-hatches on arrows using a scale from 1 to 4), much of the detail has been removed so as to focus on the pattern of connections.

Graph representations have also been developed to represent interactions between genes and proteins in a number of organisms (for pioneering work in this project, see Kauffman 1969, 1974). In developing the graph on the bottom in Fig. 9.1, Jeong et al. (2001) drew upon four datasets of possible interactions between proteins in Baker's yeast *Saccharomyces cerevisiae*. The datasets mostly relied on two-hybrid analyses to determine which proteins could interact. The nodes in this graph represent 1,870 individual proteins and the edges show 2,240 interactions (the coloring of the nodes will be discussed below).

In the sections that follow I will focus first on graphs such as these representing whole mechanisms or even the larger systems in which mechanisms of interest might be situated. In a later section I will turn to sub-graphs that show relations between components within the larger graphs. In both cases, the focus will be on how graph-theoretic tools for analyzing network organization reveal how mechanisms that instantiate the organization exhibited in the graph will behave. Although such analyses often employ mathematical models and simulations using those models, I focus on the qualitative conclusions that are reached. In the final section I show how the tools for sub-graphs can be employed after performing suitable manipulations of whole-graphs of a mechanism.

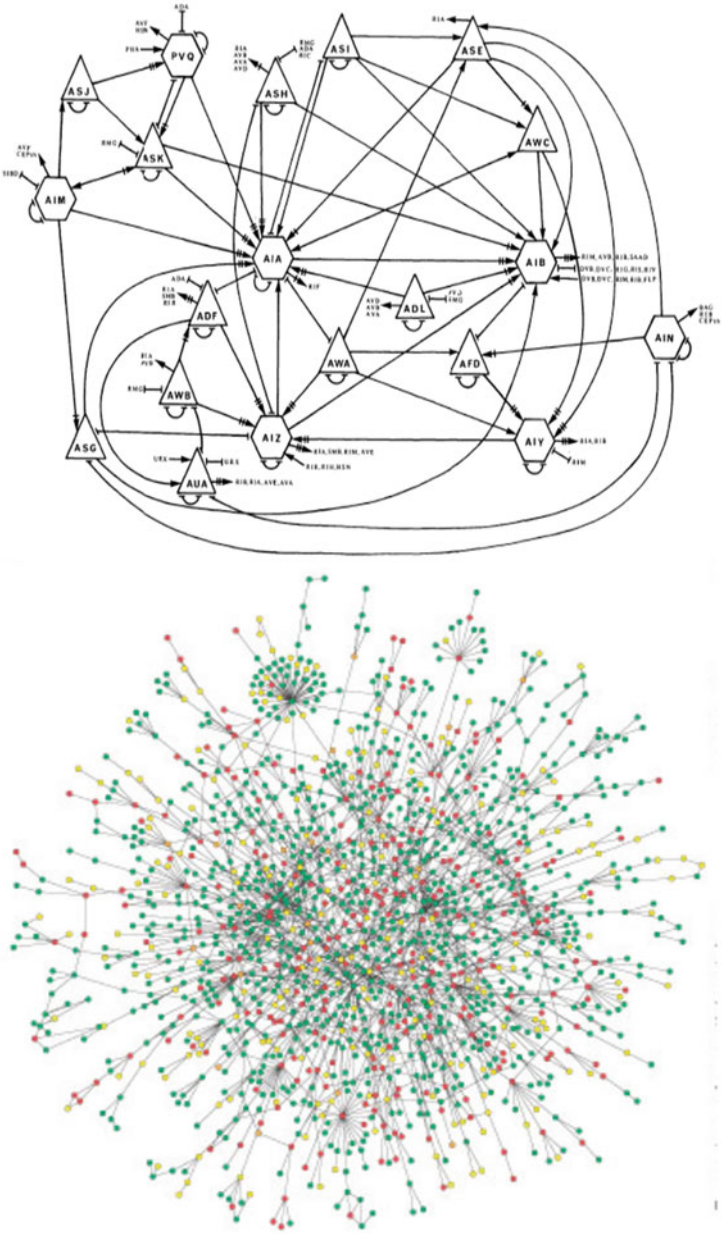

Fig. 9.1 On the *top* White et al. (1986) constructed a graph representation of that part of the nervous system of *C. elegans* involved in chemoreception. chemoreception (Reprinted from White et al., The Structure of the Nervous System of the Nematode *Caenorhabditis elegans, Proceedings of the Royal Society B, Biological Sciences*, 314, Figure 21, by permission of the Royal Society). On the *bottom* Jeong et al. (2001) proposed a graph representation of the largest cluster in the protein interaction network of *Saccharomyces cerevisiae*, containing approximately 78 % of all proteins (Reprinted by permission from Macmillan Publishers Ltd: Nature, copyright (2001))

2 Whole-Graph Organization: Scale-Free Small Worlds

Since graphs such as those shown in Fig. 9.1 are based on evidence about a particular set of entities and their connections or interactions, each is different. At first it might not be apparent that one could develop generalizable insights into how different networks will behave. Instead, one might expect that when mechanisms have sufficient parts, the graph showing the interactions will be different for each mechanism. Researchers might take advantage of the graph to develop a computational model of how a mechanism instantiating that graph might behave (Jones and Wolkenhauer 2012, demonstrate how scientists utilize graph representations as guides in developing computational models), but not be able to use them to generalize results from one mechanism to another. But pioneering graph theorists in the mid-twentieth century developed ways of analyzing the behavior of a few idealized forms of network organization that have proven very useful. A common concern for much of this work was to understand how individual entities (neurons or fireflies) that exhibit oscillatory behavior are able to synchronize their oscillations depending on how the nodes are connected. Winfree (1967) developed one of the first analyses by assuming that each of the nodes in the network is an oscillator with its own intrinsic frequency and that each is weakly connected to all the others. In his mean field model, each oscillatory node is treated as being affected by the mean activity of the other nodes. He investigated what happened as the strength of the connections between nodes is increased and demonstrated two global transitions. At first some of the oscillators begin to exhibit synchronization in their oscillation while later all of the oscillators become phase and amplitude locked. Kuramoto (1984) further advanced the analysis of mean field models.

Although the mean field approach is often used as a first approximation, a connection between each pair of nodes is not common in actual biological systems. Connections are more likely to develop between some pairs of nodes than others. Accordingly, Erdös and Rényi (1960) investigated networks in which connections are established randomly and showed that if the number of connections is much less than the number of nodes, only small, disconnected clusters of connected nodes appear in which oscillators would synchronize. However when the number of connections increases to approximately half the number of nodes, a phase transition occurs and a single giant cluster emerges in which nodes all synchronize. Even though most biological networks do not appear to exhibit random connectivity, the idea that a giant cluster will emerge at such a phase transition has been applied in analyzing networks such as protein interaction networks in yeast; the network shown on the right in Fig. 9.1 shows such a giant cluster and leaves out nodes not included in the giant cluster.[5]

[5] A special case on which extensive research has been conducted are Boolean networks (networks in which the behavior of a node is a Boolean function of the state of other nodes). See Derrida and Pomeau (1986), Flyvbjerg (1988), Kauffman (1993), Luque and Solé (1997), Rohlf and Bornholdt (2002) for analyses of Boolean networks.

Yet other graph theorists focused on networks organized as regular lattices in which nodes are connected only to their neighbors around a ring. Ermentrout and Kopell (1984) demonstrated that such networks could generate traveling waves in which oscillatory nodes around the ring would reach their maximum in succession. This approach provided a powerful tool for explaining the traveling waves exhibited in neuromuscular rhythms in the mammalian intestine as well as in the motor behavior of species such as the lamprey (Kopell and Ermentrout 1988).

Not only do random networks and regular lattices exhibit different types of behavior (synchronized activity versus traveling waves), they also exhibit patterns of connectivity that have enabled subsequent researchers to place them along a continuum. In the giant cluster of a random network, the number of edges that must be traversed on average to get from one node to another (referred to as the *characteristic path length*) is quite small whereas in regular lattices it is much greater. On the other hand, around each node in a regular lattice there is a cluster of highly interconnected nodes, whereas in a random network these clusters are much smaller. The *clustering coefficient* measures the size of these clusters. The short characteristic path length of random networks entails that activity across such a network is integrated over a relatively short period of time but the low clustering coefficient entails that there are no concentrated sets of nodes able to work together. With the high clustering in regular lattices, nodes can collaborate on a given process, but it takes much longer to relate the activity in one cluster to that occurring in others.

Watts and Strogatz (1998) explored the region between these endpoints. Starting from a regular lattice such as is shown on the left of Fig. 9.2 they investigated what happened as they replaced local connections with long-range connections and discovered that the characteristic path length dropped rapidly as the percentage of connections were replaced but that the clustering coefficient remained high until a much higher percentage was replaced. Watts and Strogatz referred to networks in this intermediate range as *small-world* networks. They argued that a host of real world networks exhibited small-world properties including not only the neural

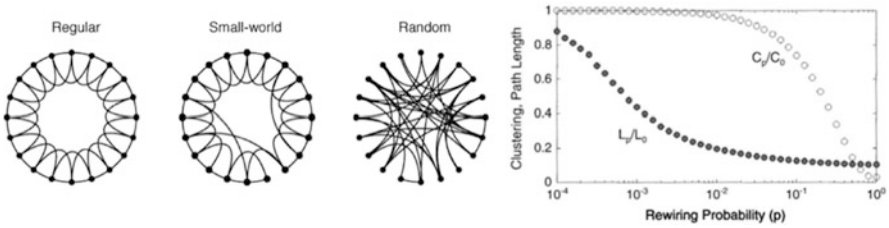

Fig. 9.2 The graphs on the *left* illustrate how small world networks are intermediate between regular lattices and random networks. One can view small-world networks as resulting from rewiring a few edges in a regular lattice. As illustrated on the *right*, as the probability of rewiring increases, the characteristic path length drops rapidly but the clustering coefficient remains high. This intermediate region is where small worlds are found. (Reprinted by permission from Macmillan Publishers Ltd: Nature, copyright (1998))

network in *C. elegans* shown in Fig. 9.1, but also the electric power grid of the Western U.S. and networks of co-appearance of actors in films. They also addressed the functional significance of small-world organization, showing for example that infectious diseases would spread much more rapidly in a small-world network than in a regular lattice. In addition, they highlighted the computational power of such networks. The key idea is that nodes that are highly clustered can be organized into appropriate sub-networks for executing particular information-processing tasks but that, as a result of the short characteristic path length, the nodes can be modulated by activity occurring elsewhere.

The two measures Watts and Strogatz introduced—characteristic path length and clustering coefficient—provide potent general measures for assessing the behavior of any systems, including mechanisms, organized in such networks. Subsequently Barabási and his colleagues addressed another measure of network architecture—node degree. This is a measure of the number of edges linking a given node. Specifically, Barabási focused on how node degree is distributed in a network. Investigations of random networks by researchers such as Erdös and Rényi assumed that node degree was distributed in a Gaussian fashion over a fairly narrow range. In such cases there is a characteristic scale at which the network can be analyzed—that is, at which component patterns can be identified. But Barabási and Albert (1999) discovered that in many real world networks, including the protein interaction network shown on the bottom in Fig. 9.1, node degree is distributed according to a power-law function so that most nodes have only a limited number of connections to other nodes while a very few have a very large number. These networks are referred to as scale-free since there is no characteristic scale at which the pattern of network activity can be identified. Barabási's initial interest seemed to be in the robustness of scale-free networks to random damage. A random removal of a node or edge would typically have little effect on the overall network. For example, mutations in the proteins shown in yellow and green in the protein interaction network in Fig. 9.1 have little or no detectable effects on the life of the yeast. These are by far most of the proteins. Those shown in red, on the other hand, are required for the cell to live while eliminating those shown in orange significantly affect growth. The red and orange nodes typically have a far higher node degree.

When node degree is distributed according to a power-law, networks tend to have modules organized around hubs. Hubs are highly connected nodes through which a large amount of the activity in a system flows. Two different types of hubs can be distinguished. Those with high clustering coefficients serve to coordinate activity within a module and are often referred to as *provincial* hub. Others, however, serve primary to link between modules, and are referred to as *connector* hubs. In many networks, including brain networks that I discuss below, there are not sufficient numbers of units to demonstrate true scale-freeness. But as long as some units have many more connections than others, hubs will likely result and will be important for understanding the behavior of the network. These networks can thus be regarded as having scale-free characteristics even if they are too small to show true scale-freeness.

I will illustrate the use of these different tools for understanding the functioning of mechanisms by focusing on research on the primate brain. Beginning with the pioneering research of Hubel and Wiesel (1962, 1968), who recorded from individual neurons in the region now known as V1 while presenting visual stimuli and demonstrated that neurons in this region responded to edges in visual scenes, researchers have identified a large number of brain regions important to visual processing, extending from the occipital lobe into the parietal and temporal cortex, and including a few frontal areas. Felleman and van Essen (1991) synthesized information about 32 cortical regions in the *Macaque* monkey involved in visual processing and the anatomical connections between them into the matrix shown on the left in Fig. 9.3. Pluses indicated the existence of a connection between the area listed on the left of each row and the area indicated at the top of each column. Visual inspection reveals clusters of areas that are interconnected. Although Felleman and van Essen did not themselves apply the measures discussed above, Sporns and Zwi (2004) calculated path length and clustering and showed the network described met the conditions for a small world. They applied similar measures and reached a similar conclusion about the matrix they developed for the whole *Macaque* brain based on data from Young (1993), shown on the right in Fig. 9.3. They also focused on sub-regions in these matrices and showed that some regions have different properties. Thus, V4 exhibits both a short path length and a low clustering coefficient, characteristic of random networks. This suggests it is a connector hub. Area A3a in the somatosensory cortex, on the other hand, is only connected to A1 and A2, which are themselves connected, giving it a long path length but high clustering. It is a provincial hub.

Techniques such as diffusion imaging and tractography have made it possible to identify in non-invasive ways interconnections of brain regions, enabling the extension of the approach to humans. Analyses of the networks identified through these techniques indicate again the existence of a small-world architecture with scale-free properties. Sporns et al. (2005) introduced the term *connectome* for these efforts at identifying and analyzing connectivity of the brain. One particularly intriguing finding from these studies which exhibits the usefulness of analyzing brain connectivity in a graph-theoretical manner was the discovery by Hagmann et al. (2008) of a set of hub regions along the anterior-posterior medial axis of the human brain which includes the rostal and caudal anterior cingulate, the paracentral lobule, and the precuneus. These regions are connected to each other and taken together connect to virtually all the other cortical areas in both hemispheres. This suggests the existence of a central network that is important for directing and coordinating information processing throughout the brain. Focusing on these areas promises to reveal how the brain coordinates specialized processing mechanisms located in different areas of the brain, enabling them to work together as needed to perform specific tasks.

The connectome project has focused primarily on anatomical connections, but a major reason for identifying anatomical connections is the assumption that these connections are likely to be important in coordinating functional activities in the brain. Bullmore and Sporns (2009) have proposed a strategy for developing

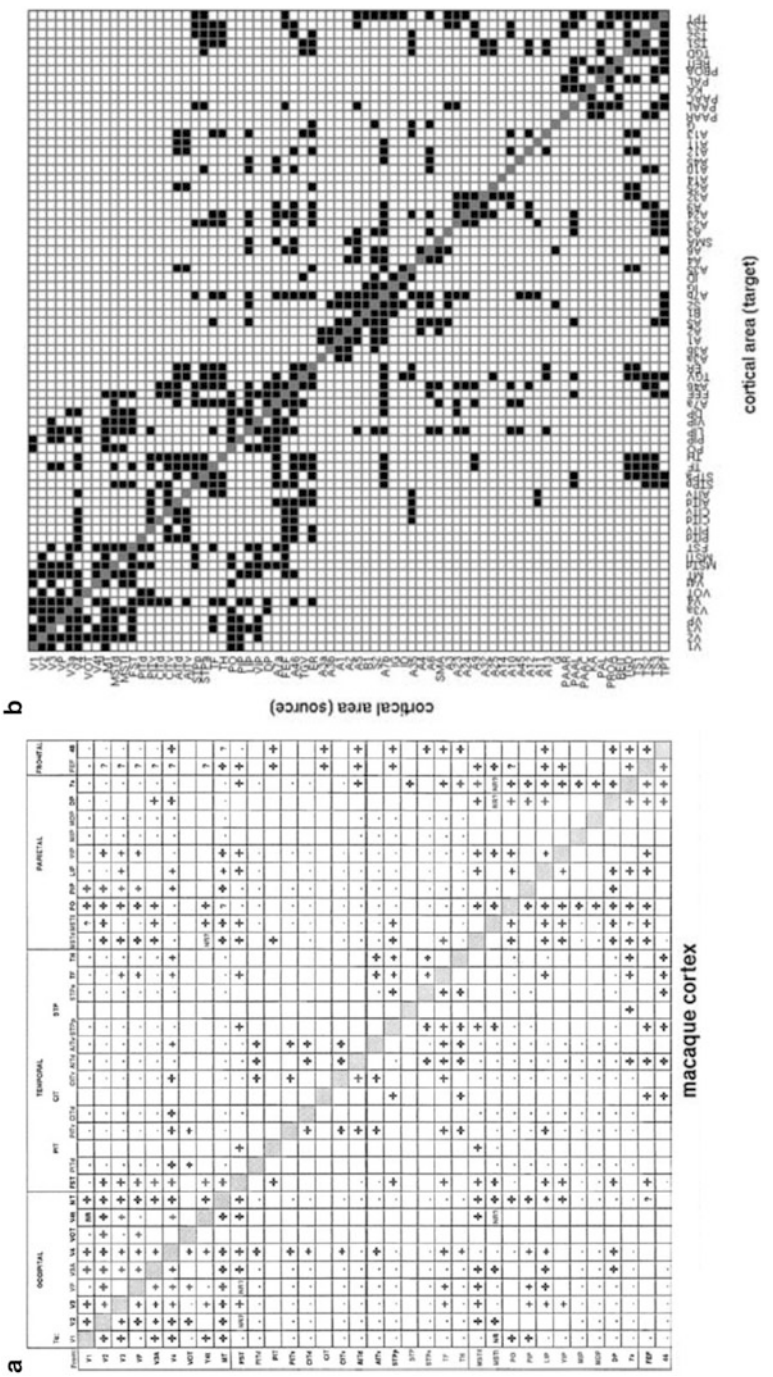

Fig. 9.3 On the *left* a matrix indicating where connections have been found between brain regions in the visual system of the *Macaque* (Felleman and van Essen, Distributed Hierarchical Processing in the Primate Cerebral Cortex, Cerebral Cortex, 1991, vol. 1, issue 1, Table 3, by permission of Oxford University Press). On the *right* is a matrix showing connectivity in the whole *Macaque* brain (Sporns and Zwi 2004, Figure 1b, with kind permission from Springer Science and Business Media)

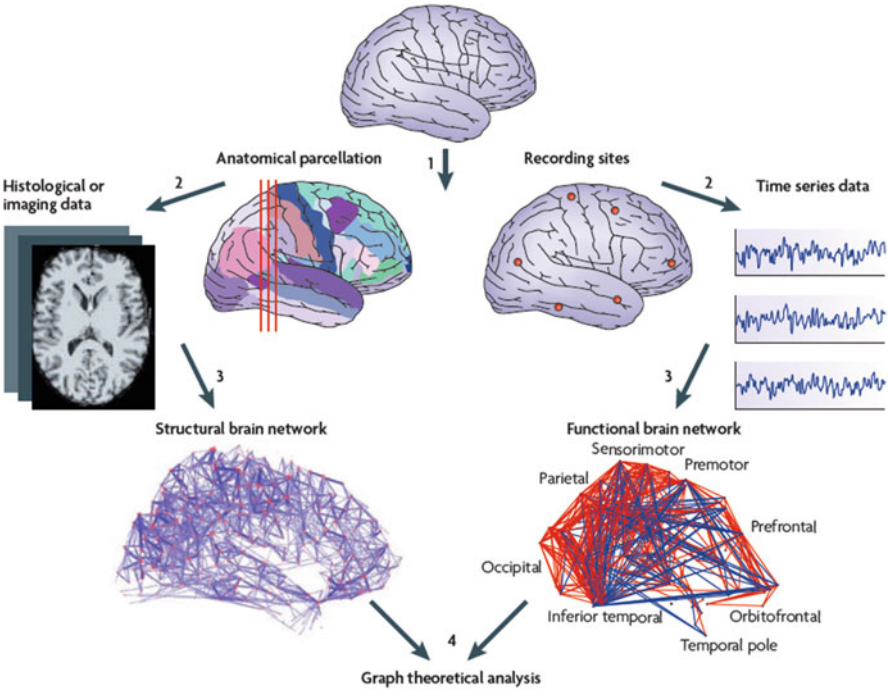

Fig. 9.4 Bullmore and Sporns (2009) proposal of a method for integrating structural and functional connectivity analyses. From each approach a graphical representation is constructed whose similarities can then be accessed. Reprinted by permission from Macmillan Publishers Ltd: Nature Reviews Neuroscience, copyright (1991)

comparable graphs of functional relations that can then be correlated to the structural graphs (see Fig. 9.4). The key to doing so is to employ techniques that provide time-series data of brain activity. EEG, which detects oscillations in electrical potentials in the 1–100 Hz range, offers one such measure. However, it is difficult to link EEG activity to specific brain regions. fMRI provides far greater localization of activity in the brain, but in initial applications was not used to support time-series analysis. When Biswal et al. (1995) did perform time series analyses on successive scans, they discovered ultraslow oscillations (<0.1 Hz) in sensorimotor areas that were correlated across hemispheres. At about the same time Raichle and his colleagues began studying people lying quietly in the scanner not performing a task (a condition referred to as the *resting state*). Shulman et al. (1997) identified several regions that exhibited greater BOLD signal in the resting state than in task conditions: the junction of precuneus and posterior cingulate cortex, the inferior parietal cortex, the left dorsolateral prefrontal cortex, a medial frontal strip that continued through the inferior anterior cingulate cortex, the left inferior frontal cortex, the left inferior frontal gyrus, and the amygdala. Since these areas were more active in the absence of a task, they interpreted them as constituting the brain's default mode network

(the network of brain regions active in non-task conditions).[6] Cordes et al. (2000) developed the technique of functional connectivity MRI (fcMRI) analysis in which they identified correlations in the oscillations in resting state BOLD signal between regions and interpreted those areas whose activity was correlated as constituting functional networks. Greicius et al. (2003) applied this approach to the default mode network and demonstrated correlated resting state activity between the regions identified as part of the default mode network. In a subsequently study Greicius et al. (2009) integrated fmMRI studies to diffusion tensor imaging of structures to establish structural connections between regions identified as parts of the default mode network, thereby implementing Bullmore and Sporns' proposed approach.

Other researchers have identified several additional networks whose activity is correlated in the resting state (Mantini et al. 2007). van den Heuvel et al. (2009) directly compared nine networks found in the resting state with diffusion tensor imaging scans and found that "well-known anatomical white matter tracks interconnect at least eight of the nine commonly found resting-state networks, including the default mode network, the core network, primary motor and visual network, and two lateralized parietal-frontal networks." Analyzing these networks graph-theoretically, He et al. (2009) found that the combined networks exhibited a small-world architecture with modules characteristic of scale-free networks, but that the individual modules exhibited different architectures that might be appropriate for the computations each performs.

In this section I have focused on graph-theoretical formalization of whole networks that represent mechanisms either individually or as constituents of larger systems such as the brain. Starting with tools to analyze random networks or regular lattices, graph theorists have developed measures such as the characteristic path length, the clustering coefficient, and the distribution of node degree. Using these measures, they have revealed small-world networks with scale-free characteristics such as nodes and hubs in many biological systems. Appeals to such features support generalizing conclusions about how mechanisms behave across different mechanisms that instantiate the same graph structure with different component parts and operations.

3 Sub-graph Organization: Motifs

In addition to ways of characterizing the organization found in whole graphs and showing the consequences different organizations have for the functioning of mechanisms exhibiting such organization, researchers have also begun to focus on sub-graphs embedded in these larger graphs that can be shown to yield specific

[6]The one type of task that yields comparable activation in these regions are episodic memory tasks, resulting in the suggestion that while resting quietly in the scanning individuals are remembering events in their lives or planning future events (see Buckner et al. 2008).

types of behavior. Until near the end of the twentieth century, only one such pattern, negative feedback, in which an edge from a subsequent node in a pathway has an inhibitory effect on one earlier in the pathway, had been subject to systematic analysis. Pioneering cyberneticists saw this as a mode of organization that could keep a process at a target level and Rosenblueth et al. (1943) argued that it was a design capable of explaining the apparent teleology or goal directness of various biological or engineered systems. While negative feedback was soon recognized as a common mode of organization in many biological systems, few biologists attended to another feature of negative feedback that physicists and engineers had identified—its capacity to support oscillations. Among the exceptions was Goodwin (1965), who analyzed negative feedback with a particular emphasis on conditions, such as non-linear interactions, that would generate sustained oscillations.

A major impetus for research on additional sub-graph structures with characteristic effects on mechanism behavior was a mode of analysis developed by Alon in his investigation of gene transcription and protein interaction networks in bacteria and yeast (for a comprehensive overview, see Alon 2007a). Figure 9.5 shows a network that represents about 20 % of the transcription interactions in *E. coli*; the nodes are genes and edges are included when the product of one gene regulates the transcription of another. Alon and his collaborators identified a number

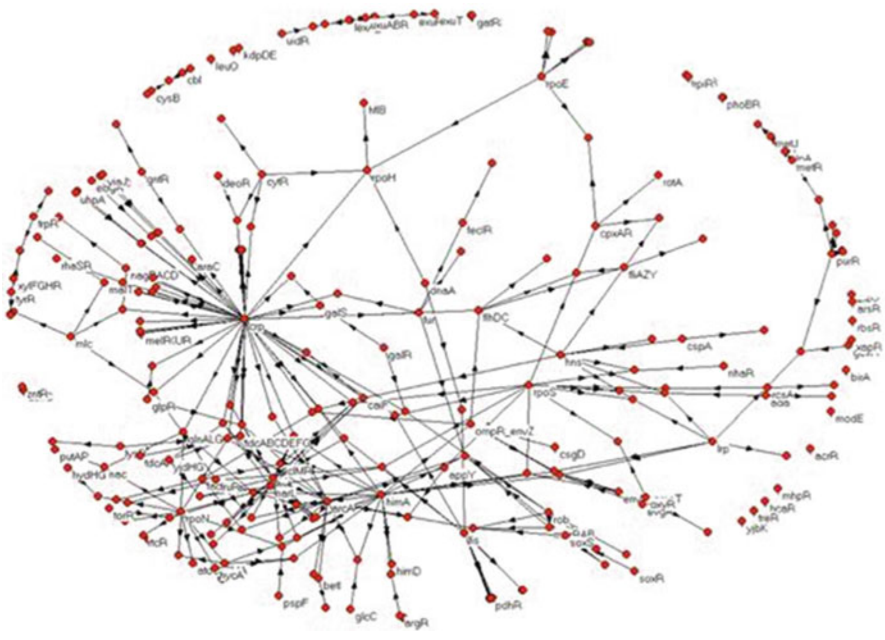

Fig. 9.5 Network representation of approximately 20 % of the transcription interactions in *E. coli*. The nodes represent genes and an edge indicates that the product of one gene exhibits regulatory effects on those of another. (Figure from Professor Uri Alon, who kindly has provided permission to reproduce it here)

of "recurring, significant patterns of interconnections" within sub-graphs (of 1–4 nodes) that appeared far more frequently than would be expected by chance (Milo et al. 2002, p. 824). He calculated the frequency expected by chance by counting how often these patterns occurred in randomly constructed networks of the same degree of node connectivity and devised an algorithm for searching databases specifying network connectivity for unusually frequently occurring subgraphs. Shen-Orr et al. (2002, p. 64) applied the term *motif* to these sub-graphs: "We generalize the notion of motifs, widely used for sequence analysis, to the level of networks. We define 'network motifs' as patterns of interconnections that recur in many different parts of a network at frequencies much higher than those found in randomized networks."[7]

Figure 9.6 illustrates three variations of a three-node sub-graph known as the *feedforward loop* that Mangan et al. (2003) reported as occurring in "hundreds of non-homologous gene systems" in the full transcription network of *E. coli*. This motif consists of three nodes representing operons in which the first (X) responds to an input signal (S) by producing a transcription factor that both regulates an operon for producing an output protein (Z) and an intermediate operation (Y) which also produces a transcription factor that regulates the output. The edges between operons can represent situations in which the first activates or inhibits the second. Making

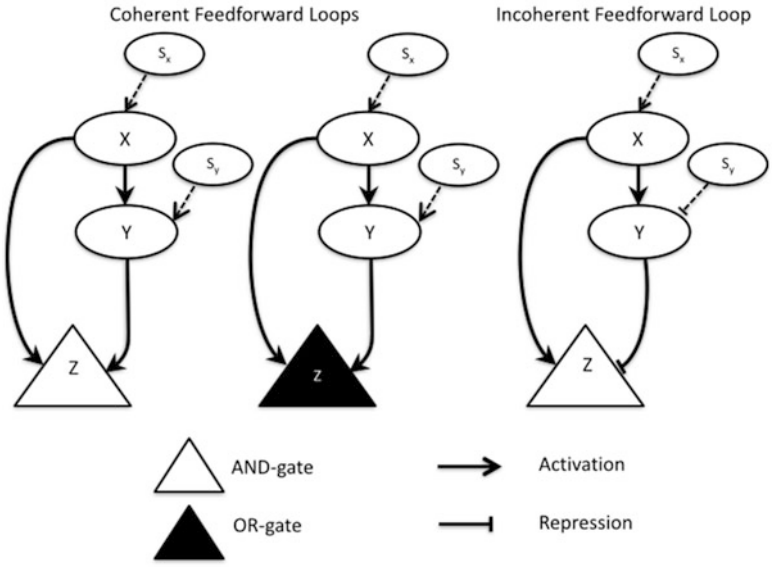

Fig. 9.6 Three versions of the feedforward loop motif investigated by Alon and his collaborators

[7]See also Braillard (2015, this volume) for a discussion of the heuristic and explanatory role of such *motifs* and more generally of modules in biology.

only very general assumptions about the component parts and operations (e.g., that each node represents a molecule that is synthesized proportional to the input that is received on the edges coming into it, that it takes time for the concentration of the molecule to build up, and that the triangular nodes operate as either AND or OR gates), Shen-Orr et al. (2002) demonstrated with a Boolean analysis that the motif on the left operates as a persistence detector. That is, it ensures that synthesis of product Z does not begin unless the input to X endures for some time. Only then will Y also accumulate so that both inputs to the AND-gate regulating Z are present. If the input to X is transient, by the time Y responds sufficiently to provide input to Z the input from X will have degraded.

By changing the AND-gate to an OR-gate, as illustrated in the middle of Fig. 9.6, the motif provides a buffer against interruption of the input to X. Keeping an AND-gate for the output operon but making the effect of Y on Z repression results in a motif that generates pulses.[8] These motifs are simple enough that one can simulate their functioning mentally. But Alon and his collaborators also provide computational simulations to support their assessment. They further identified specific instances in which the motif occurs in *E. coli* and showed how the imputed function is appropriate for controlling the particular operon occurring at Z. Thus, Mangan et al. (2003) identify the network on the left in Fig. 9.6 as occurring in the operon regulating the arabinose operon in which it is important not to commence the synthesis of the enzymes required to metabolize arabinose unless the absence of glucose, that plays the role of X, is sustained. Kalir et al. (2005) report that the motif in the middle occurs in the system regulating the construction of the flagellum where it ensures that the construction will continue until completion even if the initiating signal is interrupted. Finally, Mangan et al. (2006) identify the motif on the right as occurring in the galactose utilization system to generate a pulse of the needed enzyme galE when glucose is absent and galactose is present.

Alon and his colleagues have explored various ways in which motifs can be extended (Alon 2007b). For example, X could be replaced by two units, each of which sends inputs to Y and Z. This pattern occurs frequently in the *E. coli* transcription network. Or Z could be replaced by two or more outputs. This occurs frequently in the *C. elegans* neural network. A variant occurs in the mechanism for creating the flagellum in bacteria such as *E. coli*. The flagellum is a complex structure of nearly three-dozen parts that is built, for the most part, from the membrane out (Berg 2004; Macnab 2003). Each part must be added in sequence, which Shen-Orr et al. (2002) demonstrate can be achieved by a multi-output feedforward loop where the different outputs are regulated by parameters that set thresholds for response (see Fig. 9.7). By increasing threshold parameters K_1 to

[8]In his analysis of the neural network in *C. elegans* that he helped identify (discussed above), White (1985) identified the frequent occurrence of the feedforward loop and considered the case in which the connection from Y to Z was inhibitory. He proposed X would initially elicit a response from Z, but this would be soon be suppressed as a result of the negative connection from Y to Z. He suggests: "The whole system would therefore act as a differentiator, the output from [Z] being proportional to the rate of change of stimulus. As the animal is constantly moving, this reformation is probably of more value to it than an absolute measure of the stimulus."

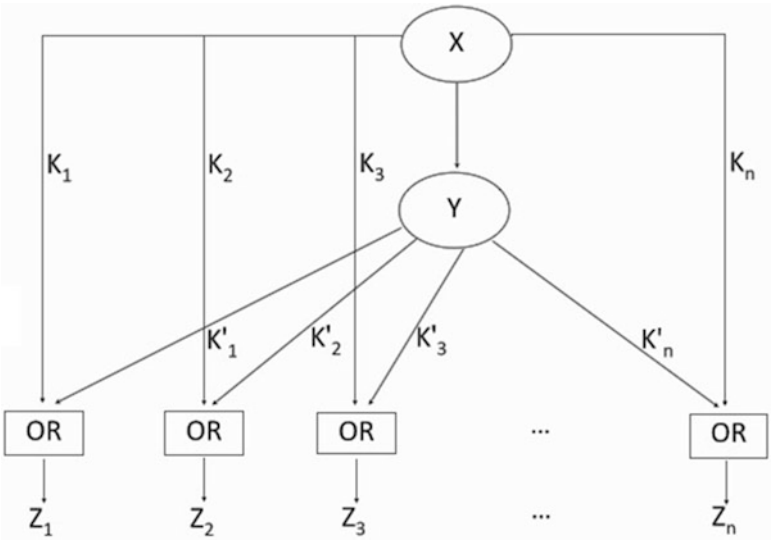

Fig. 9.7 Multi-output feedforward loop. Threshold parameters K_1 to K_n determine the order in which operons Z_1 to Z_n are activated while threshold parameters K_1' to K_n' determine the order in which they are turned off

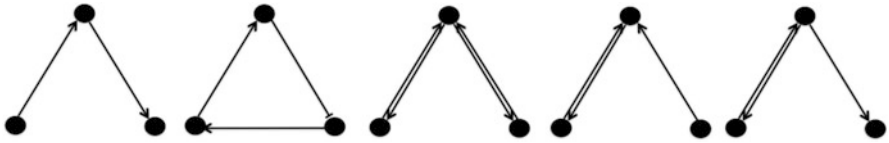

Fig. 9.8 Five different 3-node subgraphs that occur especially frequently in different networks and so qualify as motifs on Alon's approach

K_n on the connection between X and each of Z_1 to Z_n it is possible to begin each synthesis in order as the input from X increases over time (each successive Z becomes active only when the input from X has exceeded its threshold). If K_1 to K_n were the only thresholds employed, when X declined, Z_n would be the first to stop. But it is also important to stop the activity required to build the first part when it is completed. This is achieved by decreasing the parameters K_1' to K_n' on the connection between Y and each of Z_1 to Z_n. Now the first Z_n to fall below threshold will be Z_1. The resulting arrangement Shen-Orr et al. refer to refer to as *first in*, *first out*.

A key point of Alon's analysis is that different sub-graphs meet the frequency condition for being a motif in different systems. Of the 13 possible sub-graphs connecting 3 nodes, in the *E. coli* transcription network only variants of the feedforward loop occur at a frequency more than 10 standard deviations above the mean frequency found in randomly constructed networks (and just 1 of 199 possible sub-graphs connecting 4 nodes, the bi-fan, meets this condition). But in food web networks, a simple chain (shown on the top in Fig. 9.8) meets this standard, and in

electronic circuits the feedback loop (second from the left) qualifies. In their analysis of the connectivity of brain regions in the cat and monkey cortex, Sporns and Kötter (2004) report that the dual dyad (middle) occurs frequently, although its occurrence is most frequent when the apex node is a hub region.[9] In the *C. elegans* nervous system the two sub-graphs shown on the right are especially frequent.

In differentiating sub-graphs as motifs, Alon required that they occur in a given graph far more frequently then would be expected by chance. The motivation for this appears to be to provide the basis for proposing that motifs are adaptations promoted by natural selection. Several commentators have challenged use of random networks as a basis for establishing that motifs are adaptations (Artzy-Randrup et al. 2004; Solé and Valverde 2006).[10] However, one need not be an adaptationist to look to the analysis of sub-graphs to understand the design principles found in biological mechanisms (see Green et al. 2015). The analysis that Alon provides as to how sub-graphs function does not depend on motifs being adaptations or even occurring frequently. Even if a sub-graph occurs infrequently in a network, it can still perform the function revealed by mental or computational simulation of it. Accordingly, several researchers have generalized the concept of a motif to apply to small sub-graphs independently of their frequency of occurrence. Tyson and Novák (2010), for example, have developed differential equation models to explore the behavior of all two and three unit sub-graphs over a range of parameter values. A particularly interesting example from their analysis is the double-negative feedback sub-graph shown on the left in Fig. 9.9. For a range of parameter values, as illustrated on the right, this motif yields a bi-stable toggle switch—it has two stable configurations between which it can flip between but where the activation of S needed to turn on unit R is much higher than that which turns it off. In the case shown, S must be increased above $Q_{activate}$ to turn R on, but it must decrease below $Q_{inactivate}$ to turn R off. Such a sub-graph is extremely useful in situations in which it is important to maintain directionality in a process that will eventually repeat (such as the eukaryotic cell cycle) as it prevents temporary drops in the input from reversing the output but allows substantial drops to do so.

Beginning with Alon's work, researchers have analyzed sub-networks of 2–4 units and shown distinctive ways each functions by means of mental simulations and computational analyses. This provides a powerful tool for analyzing how different modes of organization found in mechanisms will affect the behavior of these mechanisms. As with whole-graph analyses, these analyses abstract from the details of the components found in the network and so, subject to certain minimal conditions, can be applied to any networks in which the sub-graphs are found.

[9]They report that computational analyses show that dual dyads are particularly effective in promoting zero phase-lag synchrony over long distances, which is important if hubs are to coordinate synchronous activity between brain regions.

[10]Ward and Thornton (2007) propose a possible origin of feedforward loops from another motif the authors term the bi-fan array to genome-wide duplication followed by random rewiring. This provides an explanation for their occurrence that doesn't depend on selection, although Ward and Thornton do not deny that selection may have also favored these motifs.

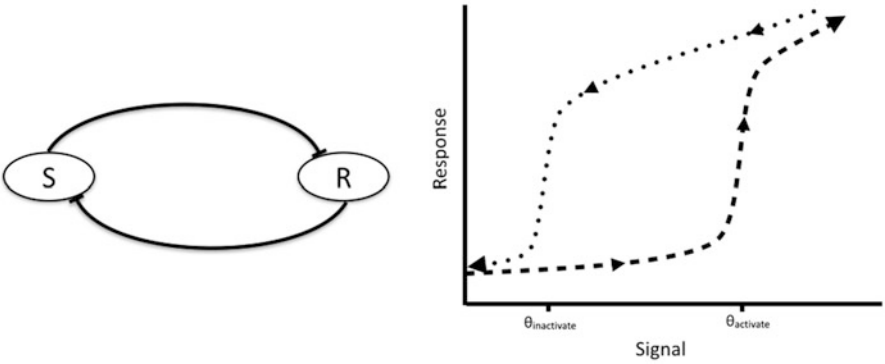

Fig. 9.9 The double negative feedback loop. The signal S and the response R each inhibit the other. With appropriate parameter values, this motif gives rise to a bi-stable switch in which the value of the signal must rise above $Q_{activate}$ for R to turn on, but once on it must decrease below $Q_{inactivate}$ before R turns off

4 Manipulating Whole-Graph Representations to Extract Motifs

In the examples discussed above, motifs were identified directly by analyzing the graphical representation of the mechanism. But in some cases investigators need to perform a series of operations on an initial graphical representation to identify motifs. I offer an example from circadian rhythm research in which such manipulations of a graphical representation have provided new insights into the working of the mechanism.

Circadian rhythms are endogenously generated oscillations of approximately 24-h that can be entrained to local cues such as light exposure and that regulate a wide range of physiological activities and behaviors. The critical mechanism is located within individual cells. As a result of the discovery of a gene (*period* or *per*) in which mutations resulted in aberrant circadian behavior in fruit flies (Konopka and Benzer 1971) and subsequent discovery that the mRNA and proteins synthesized from this gene oscillated with an approximately 24-h period, researchers hypothesized a transcriptional-translational feedback mechanism (Hardin et al. 1990). On this proposal, the protein PER synthesized from *per* mRNA feeds back to inhibit the transcription of *per* by interfering with activities at its promoter site (see the left side of Fig. 9.10). When PER finally breaks down, the inhibition is relaxed and transcription begins again. With appropriate time delays and non-linear interactions, Goldbeter (1995) showed that this mechanism could generate sustained circadian rhythms. However, subsequent research has revealed many more components to the mechanism involving multiple interacting feedback loops. The feedback loop involving *Per* is shown in the upper quadrant of the diagram on the right in Fig. 9.10 where it is now accompanied by, among other things, a positive feedback loop that produces the CLOCK:BMAL1 dimer that binds to the promoter on *Per*.

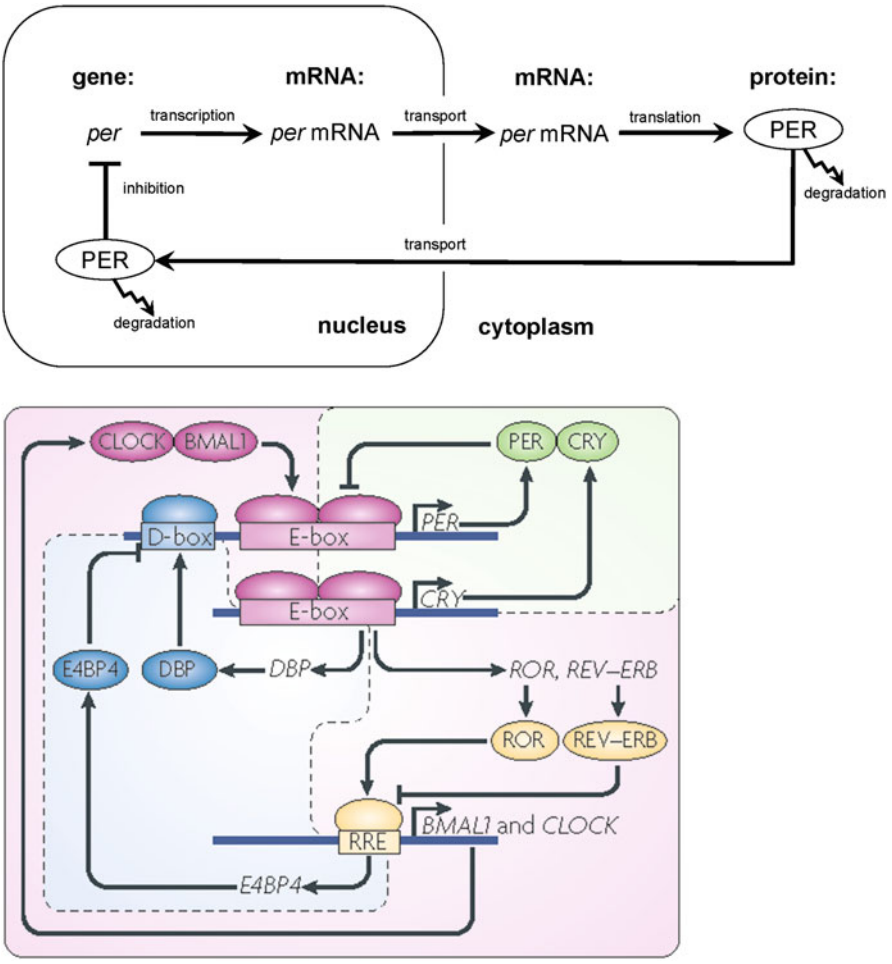

Fig. 9.10 The simple transcription-translation feedback loop as initially proposed by Hardin et al. (1990) is shown on the *top*. On the *bottom* is a representation by Zhang and Kay (2010) of the mammalian clock mechanism as it was understood circa 2005. (Reprinted by permission from Macmillan Publishers Ltd: Nature Reviews Molecular Cell Biology, copyright (2010))

A notable feature of both representations in Fig. 9.10 is that time is not explicitly represented. Rather, the viewer must follow through the sequence of operations to understand how the processes are carried out over a 24-h period. With one feedback loop, as on the *top* in Fig. 9.10, this is not too difficult, but with multiple feedback loops, as shown on the *bottom*, this becomes increasingly challenging. Activities are occurring in several feedback loops at the same time, and the figure provides no support for determining when each occurs in relation to the other. As a result, it becomes increasingly hard to understand what about the mechanism generates oscillation. In an attempt to understand the systemic organization of this

mechanism that generates oscillation, Ueda and his colleagues developed a different technique for representing the mechanism. Instead of focusing on the feedback loops themselves, they focused on what they call *clock controlled elements* (CCEs): the short DNA sequences shown near the promoter region of clock genes on the right in Fig. 9.10 and labeled E-box, D-box, and RRE. Each of these adjoins the promoter of a number of clock genes and in turn is regulated positively or negatively by another, partially overlapping set of clock genes. Using a cell-culture system in which a firefly luciferase reporter was inserted downstream of the clock-controlled promoters, Ueda et al. (2005) determined the circadian time of peak activity for each CCE: the E/E′-box regulates gene expression in the morning, the D-box regulates gene expression 5 h later in the evening, and the RRE regulates expression 8 h later during the night.

Given the distinctive timing at which each CCE regulates transcription, Ueda and his colleagues put them at the center of their graphical formalization of the mechanism (see the left side of Fig. 9.11). Time can be visualized as moving clockwise around the triangle formed by the three CCEs. The various transcription factors that regulate a CCE are shown around it—those that serve to activate a particular CCE are shown in green ovals and linked to it with green arrows while those that repress it are shown in red ovals and linked to it with bar-ended lines. There is a grey line from the node for the CCE to the transcription factors of the genes with which the CCE is associated. As an illustration, there are grey lines between the E/E′-box and the D-box to Per1 (indicating that both boxes figure in its regulation). There is also a red arrow back to the E/E′-box, indicating that Per1 represses the E/E′-boxes. This graphical formalization renders the clock genes/proteins simply as intermediaries between the three CCEs. This is further

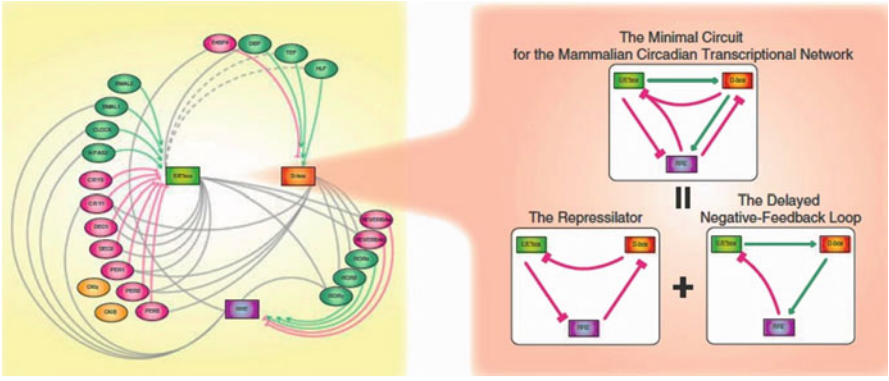

Fig. 9.11 On the *left* is Ueda's re-representation of the mammalian circadian clock mechanism in which the clock controlled elements (CCEs) are shown as *rectangles* in the center with *grey lines* linking the CCEs to the transcription factors whose genes they regulate and *green* and *red arrows* represent the effect of these transcription factors on other CCEs. On the *right* the details of the intermediary transcription factors are removed to reveal the minimal circuit shown at the *top*, which is then decomposed into two component circuits (Reprinted from Minami et al. 2013, With kind permission from Springer Science and Business Media)

illustrated by the minimal circuit diagram shown on the right in which there are excitatory and inhibitory arrows between two CCEs if there is a gene/protein associated with one CCE that has such an effect on the other CCE. For example, there is a green arrow between the E/E′-box and the D-box since the E/E-box is associated with *Dpb* and DPB is an activator of the D-box.

The point of this graphical formalization of the clock mechanism is to advance an explanation of how this mechanism is able to generate oscillatory behavior. Ueda shows that the minimal circuit can be treated as a composite of the two component circuits shown on the bottom right in Fig. 9.11—the repressilator and delayed negative feedback. Although these circuits are intended to be simplifications of complex networks, they have the same form as motifs for specific circuits. In both computational models and synthetic models, each has been shown to give rise to sustained oscillations (Elowitz and Leibler 2000; Stricker et al. 2008). When their analysis reveals the same circuit structure in these larger networks, Ueda and his colleagues conclude that the occurrence of the circuit structure in those networks explains their oscillatory behavior.

As this case indicates, researchers sometimes find it useful to modify the initial graphical formalization of a mechanism to reveal the underlying graph structural that explains the behavior of the mechanism. Thus, in this case, researchers started with a graph formalization in which all of the transcription factors were indicated (although this formalization already simplified by not distinguishing between the genes and the transcription factors). This was already an innovative formalization since it put the CCEs at the center. But the researchers then abstracted further, collapsing all the transcription factors linking one CCE to another into a single edge. This resulted in a *minimal circuit* that the researchers could further decompose into two motifs, already known from work on sub-graphs to be able to generate oscillations. The fact that they were able to reduce the whole network as formalized in a graph to two motifs known to generate oscillations enabled researchers to advance an distinctive explanation of how the much more complicated whole mechanism operates as a circadian clock. On their account, it is the fact that the relations between CCEs instantiate appropriate motifs that explains circadian rhythmicity.

5 Conclusions

A challenge in mechanistic research is how to generalize explanations developed for particular mechanisms occurring in specific tissues or organisms to mechanisms that employ similar modes of organization. Each type of mechanism (e.g., the mechanism for generating circadian rhythms in mouse SCN cells) employs its own mode of organization and this would seem to require an idiosyncratic explanation for each specific mechanism. When mechanisms employ parts that behave similarly, this knowledge can be generalized, but it has not been clear how one could do this for organization. Graph-theoretic approaches, however, are providing powerful tools

for formalizing organization patterns. By formalizing a given mechanism in a graph in which the parts are nodes and the interactions are edges, investigators abstract from the details of the specific parts and operations. Making minimal assumptions about the operations represented in edges, researchers can often draw inferences about how any mechanism exhibiting that organization will behave. Moreover, they can generalize across graphs by showing, for example, that despite differences, a set of properties is shared by the graphs that result in the same pattern of behavior. Once researchers have analyzed how mechanisms instantiating a graph structure or related graph structures behave, they can generalize that explanation to other mechanisms that instantiate the same abstract structure.

I have focused on graph theoretic analyses both of graphs representing whole mechanisms or networks of mechanisms and of sub-graphs found within larger graphs. By employing measures such as the characteristic path length, the clustering coefficient, and degree distribution, researchers have been able to classify whole-graphs and demonstrate some of the expected properties of these graphs when realized in different systems. In particular, research on scale-free small-world networks provides ways to understand both the specialized processing capacities within a larger system and the ways in which these modulate each other to achieve coordinated function. By focusing on sub-graphs of just a few nodes, researchers have demonstrated patterns of behavior different sub-graphs or motifs are expected to produce such as buffering against noise or executing operations in sequence. These can be applied to any network in which the motif is realized. I concluded with an example that showed how researchers have found ways to manipulate whole-graphs to reveal patterns of organization whose behavior is well understood. They then appealed to these organizational patterns to explain the behavior of the whole mechanism.

As biologists have discovered more and more complicated mechanisms incorporating non-linear interactions of components, the challenge of explaining their behavior has increased. Computational modeling has resulted in dynamical mechanistic analyses that attempt to integrate details about parts and operations and information about how these are organized. Graph theoretic approaches provides additional resources, enabling researchers to identify common abstract graph structures in different mechanisms and generalize the results developed for one mechanism to other mechanisms sharing that structure.

Acknowledgment Initial research on this project began when I was a Fellow at the Institute for Advanced Studies at Hebrew University. I thank the members of the group for productive discussions and especially Arnon Levy for introducing me to the work of Uri Alon and facilitating a meeting with him. Subsequently I have benefited from many further discussions with Arnon and with Sara Green. I presented much of the material here at colloquia at the University of California, Irvine and the University of Cincinnati, at workshop at the University of Wollongong, and to the reunion conference of the research group at the Institute for Advanced Studies. I thank the audiences at these various forums for very helpful comments. I also thank members of the WORGODS research group (Adele Abrahamsen, Daniel Burnston, and Benjamin Sheredos) at the University of California, San Diego for valuable discussion of the diagrammatic representations of circadian clock mechanisms. Thanks as well to Marta Halina and to the editors of this volume, Pierre-Alain Braillard and Christophe Malaterre, for valuable comments and suggestions.

References

Alon, U. (2007a). *An introduction to systems biology: Design principles of biological circuits*. Boca Raton: Chapman & Hall/CRC.

Alon, U. (2007b). Network motifs: Theory and experimental approaches. *Nature Reviews Genetics, 8*, 450–461.

Ankeny, R. A. (2001). Model organisms as models: Understanding the 'Lingua Franca' of the human genome project. *Philosophy of Science, 68*, S251–S261.

Ankeny, R. A., & Leonelli, S. (2011). What's so special about model organisms? *Studies in History and Philosophy of Science Part A, 42*, 313–323.

Artzy-Randrup, Y., Fleishman, S. J., Ben-Tal, N., & Stone, L. (2004). Comment on 'network motifs: Simple building blocks of complex networks' and 'superfamilies of evolved and designed networks'. *Science, 305*, 1107.

Baetu, T. (2015). From mechanisms to mathematical models and back to mechanisms: Quantitative mechanistic explanations. In P.-A. Braillard & C. Malaterre (Eds.), *Explanation in biology. An enquiry into the diversity of explanatory patterns in the life sciences* (pp. 345–363). Dordrecht: Springer.

Barabási, A.-L., & Albert, R. (1999). Emergence of scaling in random networks. *Science, 286*, 509–512.

Bechtel, W. (2011). Mechanism and biological explanation. *Philosophy of Science, 78*, 533–557.

Bechtel, W., & Abrahamsen, A. (2009). Decomposing, recomposing, and situating circadian mechanisms: Three tasks in developing mechanistic explanations. In H. Leitgeb & A. Hieke (Eds.), *Reduction and elimination in philosophy of mind and philosophy of neuroscience* (pp. 173–186). Frankfurt: Ontos Verlag.

Bechtel, W., & Abrahamsen, A. (2010). Dynamic mechanistic explanation: Computational modeling of circadian rhythms as an exemplar for cognitive science. *Studies in History and Philosophy of Science Part A, 41*, 321–333.

Bechtel, W., & Abrahamsen, A. (2011). Complex biological mechanisms: Cyclic, oscillatory, and autonomous. In C. A. Hooker (Ed.), *Philosophy of complex systems. Handbook of the philosophy of science* (Vol. 10, pp. 257–285). New York: Elsevier.

Bechtel, W., & Richardson, R. C. (1993/2010). *Discovering complexity: Decomposition and localization as strategies in scientific research*. Cambridge, MA: MIT Press. 1993 edition published by Princeton University Press.

Berg, H. C. (2004). *E. coli in motion*. New York: Springer.

Biswal, B., Yetkin, F. Z., Haughton, V. M., & Hyde, J. S. (1995). Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. *Magnetic Resonance in Medicine, 34*, 537–541.

Braillard, P.-A. (2015). Prospect and limits of explaining biological systems in engineering terms. In P.-A. Braillard & C. Malaterre (Eds.), *Explanation in biology. An enquiry into the diversity of explanatory patterns in the life sciences* (pp. 319–344). Dordrecht: Springer.

Brigandt, I. (2013). Systems biology and the integration of mechanistic explanation and mathematical explanation. *Studies in History and Philosophy of Biological and Biomedical Sciences, 44*, 477–492.

Buckner, R. L., Andrews-Hanna, J. R., & Schacter, D. L. (2008). The brain's default network: Anatomy, function, and relevance to disease. *Annals of the New York Academy of Sciences, 1124*, 1–38.

Bullmore, E., & Sporns, O. (2009). Complex brain networks: Graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience, 10*, 186–198.

Cordes, D., Haughton, V. M., Arfanakis, K., Wendt, G. J., Turski, P. A., Moritz, C. H., Quigley, M. A., & Meyerand, M. E. (2000). Mapping functionally related regions of brain with functional connectivity MR imaging. *American Journal of Neuroradiology, 21*, 1636–1644.

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

Derrida, B., & Pomeau, Y. (1986). Random networks of automata: A simple annealed approximation. *Europhysics Letters, 1*, 45–49.

Elowitz, M. B., & Leibler, S. (2000). A synthetic oscillatory network of transcriptional regulators. *Nature, 403*, 335–338.

Erdös, P., & Rényi, A. (1960). On the evolution of random graphs. *Proceedings of the Mathematical Institute of the Hungarian Academy of Sciences, 5*, 17–61.

Ermentrout, G. B., & Kopell, N. (1984). Frequency plateaus in a chain of weakly coupled oscillators. 1. *Siam Journal on Mathematical Analysis, 15*, 215–237.

Felleman, D. J., & van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex, 1*, 1–47.

Flyvbjerg, H. (1988). An order parameter for networks of automata. *Journal of Physics A: Mathematical and General, 21*, L955–L960.

Glennan, S. (1996). Mechanisms and the nature of causation. *Erkenntnis, 44*, 50–71.

Glennan, S. (2002). Rethinking mechanistic explanation. *Philosophy of Science, 69*, S342–S353.

Goldbeter, A. (1995). A model for circadian oscillations in the *Drosophila* period protein (PER). *Proceedings of the Royal Society of London B: Biological Sciences, 261*, 319–324.

Goodwin, B. C. (1965). Oscillatory behavior in enzymatic control processes. *Advances in Enzyme Regulation, 3*, 425–428.

Green, S., Levy, A., & Bechtel, W. (2015). Design sans adaptation. *European Journal for the Philosophy of Science, 5*, 15–29.

Greicius, M. D., Krasnow, B., Reiss, A. L., & Menon, V. (2003). Functional connectivity in the resting brain: A network analysis of the default mode hypothesis. *Proceedings of the National Academy of Sciences of the United States of America, 100*, 253–258.

Greicius, M. D., Supekar, K., Menon, V., & Dougherty, R. F. (2009). Resting-state functional connectivity reflects structural connectivity in the default mode network. *Cerebral Cortex, 19*, 72–78.

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*, e159.

Hardin, P. E., Hall, J. C., & Rosbash, M. (1990). Feedback of the *Drosophila period* gene product on circadian cycling of its messenger RNA levels. *Nature, 343*, 536–540.

He, Y., Wang, J., Wang, L., Chen, Z. J., Yan, C., Yang, H., Tang, H., Zhu, C., Gong, Q., Zang, Y., & Evans, A. C. (2009). Uncovering intrinsic modular organization of spontaneous brain activity in humans. *PLoS One, 4*, e5226.

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiology, 160*, 106–154.

Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *Journal of Physiology, 195*, 215–243.

Issad, T., & Malaterre, C. (2015). Are dynamic mechanistic explanations still mechanistic? In P.-A. Braillard & C. Malaterre (Eds.), *Explanation in biology. An enquiry into the diversity of explanatory patterns in the life sciences* (pp. 265–292). Dordrecht: Springer.

Jeong, H., Mason, S. P., Barabasi, A. L., & Oltvai, Z. N. (2001). Lethality and centrality in protein networks. *Nature, 411*, 41–42.

Jones, N. (2014). Bowtie structures, pathway diagrams, and topological explanation. *Erkenntnis, 79*, 1135–1155.

Jones, N., & Wolkenhauer, O. (2012). Diagrams as locality aids for explanation and model construction in cell biology. *Biology and Philosophy, 27*, 705–721.

Kalir, S., Mangan, S., & Alon, U. (2005). A coherent feed-forward loop with a SUM input function prolongs flagella expression in *Escherichia coli. Molecular Systems Biology, 1*, 2005.0006.

Kauffman, S. A. (1969). Metabolic stability and epigenesis in randomly constructed genetic nets. *Journal of Theoretical Biology, 22*(3), 437–467.

Kauffman, S. A. (1974). The large scale structure and dynamics of gene control circuits: An ensemble approach. *Journal of Theoretical Biology, 44*(1), 167–190.

Kauffman, S. A. (1993). *The origins of order: Self-organization and selection in evolution*. New York: Oxford University Press.

Konopka, R. J., & Benzer, S. (1971). Clock mutants of *Drosophila melanogaster. Proceedings of the National Academy of Sciences of the United States of America, 89*, 2112–2116.

Kopell, N., & Ermentrout, G. B. (1988). Coupled oscillators and the design of central pattern generators. *Mathematical Biosciences, 90*, 87–109.

Kuramoto, Y. (1984). *Chemical oscillations, waves, and turbulence*. Berlin: Springer.

Leonelli, S., Ramsden, E., Nelson, N., & Ankeny, R. A. (2014). Making organisms model humans: Situated models in alcohol research. *Science in Context, 27*(3), 485–509.

Levy, A., & Bechtel, W. (2013). Abstraction and the organization of mechanisms. *Philosophy of Science, 80*, 241–261.

Luque, B., & Solé, R. (1997). Phase transitions in random networks: Simple analytic determination of critical points. *Physical Review E, 55*, 257–260.

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

Macnab, R. M. (2003). How bacteria assemble flagella. *Annual Review of Microbiology, 57*, 77–100.

Mangan, S., Zaslaver, A., & Alon, U. (2003). The coherent feedforward loop serves as a sign-sensitive delay element in transcription networks. *Journal of Molecular Biology, 334*, 197–204.

Mangan, S., Itzkovitz, S., Zaslaver, A., & Alon, U. (2006). The incoherent feed-forward loop accelerates the response-time of the gal system of *Escherichia coli. Journal of Molecular Biology, 356*, 1073–1081.

Mantini, D., Perrucci, M. G., Del Gratta, C., Romani, G. L., & Corbetta, M. (2007). Electrophysiological signatures of resting state networks in the human brain. *Proceedings of the National Academy of Sciences, 104*, 13170–13175.

Meunier, R. (2012). Stages in the development of a model organism as a platform for mechanistic models in developmental biology: Zebrafish, 1970–2000. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences, 43*, 522–531.

Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., & Alon, U. (2002). Network motifs: Simple building blocks of complex networks. *Science, 298*, 824–827.

Minami, Y., Ode, K. L., & Ueda, H. R. (2013). Mammalian circadian clock: The roles of transcriptional repression and delay. *Handbook of Experimental Pharmacology, 217*, 359–377.

Overton, J. A. (2011). Mechanisms, types, and abstractions. *Philosophy of Science, 78*, 941–954.

Rohlf, T., & Bornholdt, S. (2002). Criticality in random threshold networks: Annealed approximation and beyond. *Physica A: Statistical Mechanics and its Applications, 310*, 245–259.

Rosenblueth, A., Wiener, N., & Bigelow, J. (1943). Behavior, purpose, and teleology. *Philosophy of Science, 10*, 18–24.

Shen-Orr, S. S., Milo, R., Mangan, S., & Alon, U. (2002). Network motifs in the transcriptional regulation network of Escherichia coli. *Nature Genetics, 31*, 64–68.

Shulman, G. L., Corbetta, M., Buckner, R. L., Fiez, J. A., Miezin, F. M., Raichle, M. E., & Petersen, S. E. (1997). Common blood flow changes across visual tasks: I. increases in subcortical structures and cerebellum but not in nonvisual cortex. *Journal of Cognitive Neuroscience, 9*, 624–647.

Solé, R. V., & Valverde, S. (2006). Are network motifs the spandrels of cellular complexity? *Trends in Ecology and Evolution, 21*, 419–422.

Sporns, O., & Kötter, R. (2004). Motifs in brain networks. *PLoS Biology, 2*, e369.

Sporns, O., & Zwi, J. D. (2004). The small world of the cerebral cortex. *Neuroinformatics, 2*, 145–162.

Sporns, O., Tononi, G., & Kötter, R. (2005). The human connectome: A structural description of the human brain. *PLoS Computational Biology, 1*, e42.

Stricker, J., Cookson, S., Bennett, M. R., Mather, W. H., Tsimring, L. S., & Hasty, J. (2008). A fast, robust and tunable synthetic gene oscillator. *Nature, 456*, 516–519.

Théry, F. (2015). Explaining in contemporary molecular biology: Beyond mechanisms. In P.-A. Braillard & C. Malaterre (Eds.), *Explanation in biology. An enquiry into the diversity of explanatory patterns in the life sciences* (pp. 113–133). Dordrecht: Springer.

Tyson, J. J., & Novák, B. (2010). Functional motifs in biochemical reaction networks. *Annual Review of Physical Chemistry, 61*, 219–240.

Ueda, H. R., Hayashi, S., Chen, W., Sano, M., Machida, M., Shigeyoshi, Y., Iino, M., & Hashimoto, S. (2005). System-level identification of transcriptional circuits underlying mammalian circadian clocks. *Nature Genetics, 37*, 187–192.

van den Heuvel, M. P., Mandl, R. C. W., Kahn, R. S., & Pol, H. E. H. (2009). Functionally linked resting-state networks reflect the underlying structural connectivity architecture of the human brain. *Human Brain Mapping, 30*, 3127–3141.

Ward, J. J., & Thornton, J. M. (2007). Evolutionary models for formation of metwork motifs and modularity in the *Saccharomyces* transcription factor network. *PLoS Computational Biology, 3*, e198.

Watts, D., & Strogratz, S. (1998). Collective dynamics of small worlds. *Nature, 393*, 440–442.

White, J. G. (1985). Neuronal connectivity in Caenorhabditis elegans. *Trends in Neurosciences, 8*, 277–283.

White, J. G., Southgate, E., Thomson, J. N., & Brenner, S. (1986). The structure of the nervous system of the nematode *Caenorhabditis elegans. Philosophical Transactions of the Royal Society of London. B Biological Sciences, 314*, 1–340.

Winfree, A. T. (1967). Biological rhythms and the behavior of populations of coupled oscillators. *Journal of Theoretical Biology, 16*, 15–42.

Woodward, J. (2013). II—Mechanistic explanation: Its scope and limits. *Aristotelian Society Supplementary Volume, 87*, 39–65.

Young, M. P. (1993). The organization of neural systems in the primate cerebral cortex. *Proceedings of the Royal Society of London. Series B: Biological Sciences, 252*, 13–18.

Zhang, E. E., & Kay, S. A. (2010). Clocks not winding down: Unravelling circadian networks. *Nature Reviews Molecular and Cell Biology, 11*, 764–776.