Roles of Diagrams in Computational Modeling of Mechanisms

William Bechtel (bechtel@ucsd.edu)
Department of Philosophy and Center for Chronobiology, University of California,
San Diego, La Jolla, CA 92093-0119 USA

Adele Abrahamsen (aabrahamsen@ucsd.edu)
Center for Research in Language, University of California, San Diego
La Jolla, CA 92093-0119 USA

Abstract

As tools in science, diagrams not only serve as vehicles for communication but also facilitate and constrain scientific reasoning. We identify roles that diagrams play when computational models and synthesized organisms are used to recompose mechanisms proposed to explain biological phenomena. Diagrams not only serve as locality aids for constructing computational models but also help in identifying ways to manipulate these models and interpret the results. Moreover, they serve as blueprints for constructing synthetic organisms and then guide the interpretation of discrepancies between these organisms and computational models.

Keywords: diagrams; computational models; mechanistic explanation; circadian rhythms

Introduction

Cognitive scientists have contributed analyses and experiments on the roles diagrams play in reasoning and problem solving (e.g., Hegarty, 2004, 2011; Tversky, 2011) and have even designed new diagram formats that facilitate learning in math and science (Cheng, 2002, 2011). However, there have been only a few studies of the roles diagrams play in the natural sciences (Nersessian, 2008; Gooding, 2010). The most obvious role, evidenced by the ubiquity of diagrams in talks and publications, is communication of methods, results, and proposed mechanistic explanations (Perini, 2005). Less visibly, but crucially, diagramming is a tool that scientists use to reason about phenomena (Bechtel & Abrahamsen, 2012) and the mechanisms that might explain them (Sheredos, Burnston, Abrahamsen, & Bechtel, in press).

In many fields of biology, such as cell and molecular biology, the primary goal of research in the 19th and 20th centuries was to identify and *decompose* mechanisms to determine their parts (e.g., proteins) and operations (e.g., catalyzing particular chemical reactions). As recognized in the new mechanistic philosophy of science, the organization of these parts and operations must also be determined to arrive at a basic mechanistic explanation of a phenomenon of interest (Bechtel & Richardson, 1993/2010; Bechtel & Abrahamsen, 2005; Machamer, Darden, & Craver, 2000; Thagard, 2003). That is, to understand how the parts and operations contribute to producing the phenomenon, researchers must *recompose* the responsible mechanism either conceptually or physically. Through most of the 20th century this involved proposing a simple sequence in which the operations might occur, perhaps using mental simulation to verify its plausi-

bility (Bechtel, 2006). By the last decades of the century, however, the operations of numerous biological mechanisms were understood to display nonlinear, continuous dynamics and complex interactions. As sequential organization broke down, so too did biologists' ability to mentally track the functioning of the proposed mechanisms. Hence, they turned first to computational models and later to synthetic organisms as tools for recomposing mechanisms, with an emphasis on investigating the complex dynamics and interactions of operations by which a mechanism generates a phenomenon. In this paper we identify some of the roles diagrams play in the design of computational and synthetic models of mechanisms in actual scientific practice.

Computational modeling in biology, in contrast to that in much of cognitive science, has been grounded in considerable knowledge of the physical parts and operations of the mechanisms being targeted (Bechtel & Abrahamsen, 2010). Diagrams showing how different parts are thought to operate on each other serve as *locality aids* that "group together information that is used together" in the mechanism itself and hence often in computational models of its dynamics (Jones & Wolkenhauer, 2012, p. 705). But such diagrams also figure centrally in conceiving how manipulations made to the computational model correspond to possible perturbations of the mechanism, thereby relating experiments on models to experiments on actual mechanisms or to pathologies known to result from damage to actual mechanisms. Moreover, as the efforts to recompose mechanisms increasingly take a step beyond computational modeling to synthesizing organisms, a diagram can serve both as a blueprint for synthesizing an organism and as a medium for adjudicating mismatches in behavior between organism and model.

We focus on one domain of biology, circadian rhythms: the daily oscillations in a variety of physiological and behavioral processes in species ranging from bacteria and fungi to plants and animals. The phenomena of greatest interest involve three characteristics of these rhythms: they are endogenously generated, entrained to the day-night cycle on our planet, and sustained over time (not dampened). Their complex dynamics have made circadian rhythm research a model case for developing computational models and synthesized organisms to determine how a proposed mechanism might account for relevant phenomena. By examining specific exemplars of this research, we show how diagrams can play an important role in the reasoning that goes into computational modeling and synthetic biology.

Diagrams for Modeling "How-Possibly" Mechanisms

Computational modeling of circadian rhythms began shortly after behavioral researchers determined that the daily oscillations in organisms are endogenously generated, with a period varying slightly from 24 hours (Bünning, 1960). Since engineers had shown that negative feedback systems generate oscillations, biologists were attracted to the idea that feedback loops are involved in circadian oscillations. But most feedback systems dampen, settling into a steady state. The challenge was to determine how a biological mechanism might generate sustained oscillations, which entailed computational modeling of its dynamics.

Goodwin (1963) accepted this challenge, and took as his starting point one of the first molecular feedback mechanisms identified in biology: the lac operon. Jacob and Monod (1961) had specified how synthesis of the enzymes needed to metabolize lactose could be restricted via negative feedback to occur only when glucose levels are low. Although the molecular parts and operations involved in the circadian mechanism had not yet been identified, Goodwin borrowed the architecture of the better-understood lac operon to construct a diagram depicting a possible circadian mechanism (Figure 1). In it he included not only generic labels for the putative parts and operations but also associated variables and parameters relevant to their dynamics. The mechanism has five types of molecular parts, three of which undergo changes in their concentration. These concentrations are represented by the variables X, Y, and Z. Arrows depict six operations that affect the concentrations: three (labeled) involve aspects of gene expression and three indicate decay of a particular type of molecule, at rates associated with the parameters $k_1, \ldots k_6$. Thus, X is the concentration of mRNA transcribed from the gene, Y the concentration of the enzyme resulting from translating the mRNA, Z the concentration of the repressor molecule whose synthesis is catalyzed by the enzyme, k_4 to k_6 the rates of decay, and k_1, to k_3 associated with rates of gene expression operations.

There are three equations in the computational model. Each specifies the change in concentration of one molecular component by subtracting a term for its decay from a term for the impact of one of the operations in the feedback loop. Consulting the diagram, it is easy to see which variables and parameters should be in the same equation. Each variable has one arrow from it (its decay) and one arrow to it from another variable; its equation includes that variable and the parameters on those arrows. By providing these groupings, the diagram does service as a locality aid.

$$\frac{dX}{dt} = \frac{k_1}{Z^n + 1} - k_4 X$$

$$\frac{dY}{dt} = k_2 X - k_5 Y$$

$$\frac{dZ}{dt} = k_3 Y - k_6 Z$$

Five of the terms simply multiply a concentration by a rate parameter. The first term is more complex: since the repressor *reduces* synthesis of mRNA, its concentration (Z) is in the denominator and raised to the power n; known as the Hill coefficient, n represents the number of molecules that must interact. As the only nonlinear term, this first term is crucial for generating sustained oscillations.

It is difficult to determine exactly how a mechanism will behave when even one component exhibits nonlinearity and also when appropriate parameter values are not yet known. For both of these reasons, it is important to run simulations by solving the equations with different initial values and parameter settings. Doing so on an analog computer, Goodwin concluded that such a mechanism could generate sustained oscillations when n equaled 2 or 3. These are biologically plausible values, but when Griffith (1968) ran simulations on a digital computer he determined that sustained oscillations resulted only when $n>9$, generally recognized as biologically unrealistic. Accordingly, he concluded that negative feedback with a single gene product operating on a gene could never "give rise in practice to undamped oscillations in the concentrations of cellular constituents" (p. 207). This reasoning highlights an advantage of grounding a computational model in a representation of the associated mechanism. A biologist, having noticed that the term in question relates to molecules interacting to inhibit a biochemical reaction, can draw on knowledge of such reactions to judge the plausibility of different parameter values. Lacking such grounding, the modeler has no independent check on the values obtained from parameter fitting.
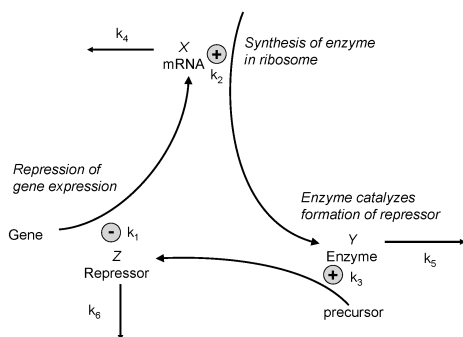

Figure 1. Diagram of the generic mechanism for feedback control of gene expression that Goodwin used as a locality aid in constructing his computational model of circadian rhythms (adapted from Goodwin 1963).

Diagrams for Modeling Known Parts and Operations

Diagrams continued to serve as locality aids after researchers discovered some of the actual parts and operations of the circadian mechanism, and modelers turned to modeling their specific dynamics. As we will see, the diagrams also supported additional reasoning about the mechanism.

The first component part of a circadian clock was discovered by Konopka and Benzer (1971) through a process of generating mutant fruit flies with short, long, or absent circadian rhythms. They named the gene in which mutations

produced altered rhythms *period* (*per*). When cloning techniques became available, Hardin, Hall, and Rosbash (1990) were able to measure the mRNA into which *per* was transcribed and the protein into which it was translated. They determined that these concentrations oscillated over 24 hours, with the peak concentration of the protein lagging several hours behind that of the mRNA transcript. They thus hypothesized a feedback mechanism whereby the protein PER fed back to inhibit the transcription of the gene *per*.

This research physically identified some of the parts and operations of the proposed mechanism, but the "feedback hypothesis" left open the question of whether and under what specific conditions it could generate sustained oscillations. Goldbeter (1995) took up this question by developing a computational model, drawing upon Hardin et al.'s empirical discoveries and inspired in part by Goodwin's abstract model. Like Goodwin, he portrayed the mechanism in a diagram (Figure 2) in which each part and operation was accompanied by its corresponding variable or parameter. Shown within the dashed box is the operation occurring in the nucleus in which the PER protein inhibits *per* transcription. The rest of the diagram shows the operations of transcription and translation and an additional post-translational operation through which the protein PER is phosphorylated
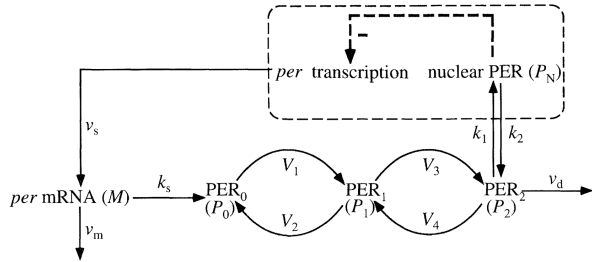
(a step that had been determined to be necessary before PER could be transported back into the nucleus).

Like Goodwin, Goldbeter then constructed differential equations, each characterizing the change in concentration of one of the molecular components. Again, the grouping of arrows around each variable served as a locality aid in determining the equations. As a result of including additional nonlinearities in the terms representing decay, Goldbeter's model exhibited sustained oscillations using parameter values deemed biologically realistic.

In the same window of time during which Goldbeter was constructing his model, molecular researchers were searching for additional parts to fill known gaps in the mechanism. They recognized, for example, that PER could not directly inhibit its own transcription since it lacked the needed binding region. Mammalian researchers identified a gene, *Clock*, in which a mutation could eliminate circadian function and whose protein contained a DNA-binding region (Vitaterna, King, Chang, Kornhauser, Lowrey, McDonald, Dove, Pinto, Turek, & Takahashi, 1994). In short order, it was found that CLOCK forms a dimer with BMAL1 that binds to the promoter region of *Per* (as well as a second gene, *Cry*) and that by interacting with this dimer, PER and CRY inhibit their own transcription. Realizing that concentrations of BMAL1 oscillate, researchers hypothesized a second negative feedback loop in which it inhibited the transcription of its gene. The introduction of this additional feedback loop raised the question of whether the results of Goldbeter's (1995) simulation were still applicable: would the two loops generate sustained oscillations? To address this question, Leloup and Goldbeter (2003) constructed a diagram (Figure 3) that included a variable for the concentration of each molecular part and a rate parameter for each operation. Again, the grouping of arrows around each variable served as a locality aid. With 16 variables being tracked this time, the computational model consisted of 16 differential equations.
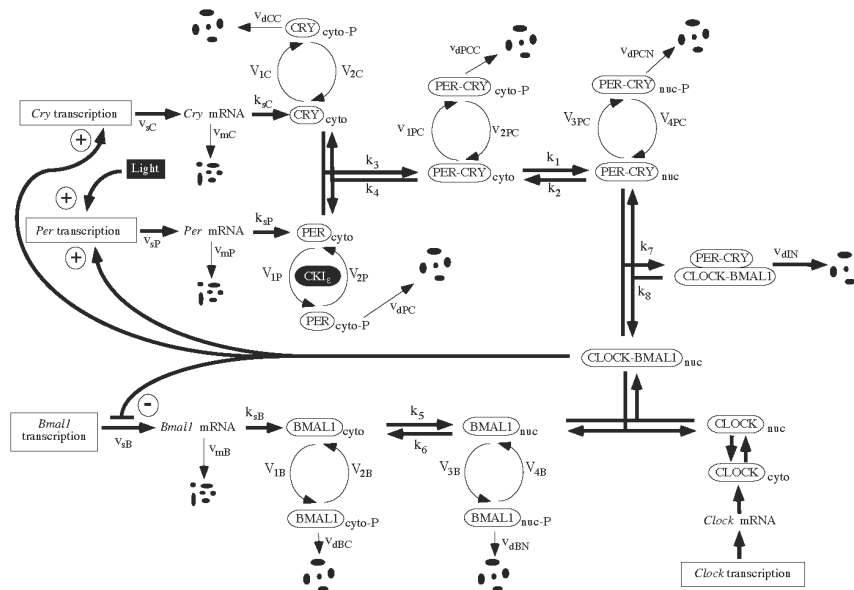

Figure 2. Goldbeter's (1995) diagram that guided his computational model based on the mechanism proposed by Hardin, Hall, and Rosbash (1990).

Figure 3. Leloup and Goldbeter's (2004) diagram of the mammalian circadian oscillator in which proteins are represented as ovals (labeled within) and operations as arrows (some identified in adjacent boxes, and all with rate parameters shown).

Leloup and Goldbeter employed their computational model not only to establish that the mechanism could generate sustained oscillations, but to determine as well whether it could account for other circadian phenomena. Of prime importance is the ability of circadian clocks to be entrained by light. Light had been shown experimentally to affect PER expression, and hence Leloup and Goldbeter incorporated light in their diagrams as a black box with an arrow feeding into the box for *Per* transcription. This in turn guided their strategy for simulating light exposure in the computational model: instead of a setting a single value for the parameter v_{sP}, which set the maximum rate of *Per* expression, they used a square wave function to alternate between a high value (simulating light) and a low value (simulating darkness). Leloup and Goldbeter were then able to use their model to show that the mechanism's responses to light exposure varied with time of day in ways similar to the responses of mammals.

Leloup and Goldbeter were also interested in whether the proposed mechanism could be perturbed in ways that correspond to known circadian pathologies. Advanced sleep phase syndrome is a condition in which people naturally go to sleep around 7 PM and rise around 3 AM. Genetic studies of families with this pathology had revealed a mutation affecting the interaction of PER with a kinase that phosphorylates it. The diagram includes the parameter v_{1P} at this location, and Leloup and Goldbeter showed that they could replicate the characteristics of the pathology by altering it.

In a subsequent paper Leloup and Goldbeter (2004) explored the sensitivity of the model to variations in all of the parameters. Here the diagram facilitated identifying which operations in the actual mechanism correspond to those perturbed by varying parameters in the computational model.

A question researchers often ask when they encounter a mechanistic account is whether all of the parts are required for the phenomenon to occur. Leloup and Goldbeter questioned which of the two feedback loops in their diagram were essential for circadian rhythmicity, and explored this by setting the parameter governing PER synthesis to 0. The model ceased to exhibit oscillation. They then explored whether oscillation could be rescued by increasing parameters regulating the synthesis of BMAL1. This restored oscillation, but with a shorter period of approximately 19 hours.

This question of what different components contribute to the generation of circadian rhythms remains one of great interest to modelers. Some have pursued the question using highly reduced models, but adopting Goldbeter's approach instead, Relógio, Westermark, Wallach, Schellenberg, Kramer, & Herzel (2011) included in their model all of the currently identified operations in the mammalian circadian mechanism. They developed the diagram in Figure 4 as a locality aid. Like the other diagrams, it includes variables and parameters adjacent to the relevant parts and operations. An innovation is use of a dashed line to differentiate two sub-mechanisms. By running the model with targeted variables set to constant values—first those for concentrations of parts above the line and then those below—they concluded

that it was the feedback loops involving BMAL1 that were crucial to the generation of circadian rhythms.
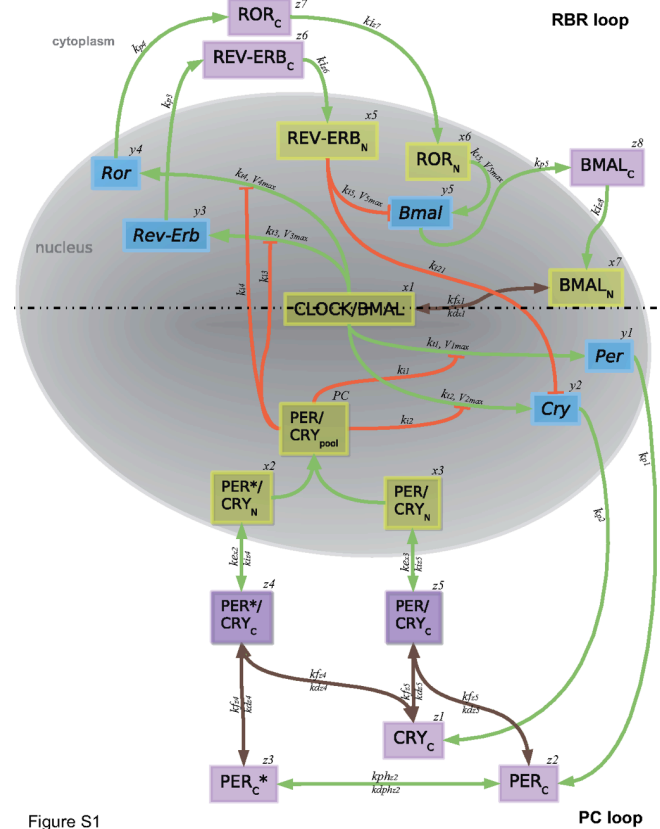

Figure S1

Figure 4. Relógio et al's (2011) diagram of the mammalian circadian oscillator. They use a dotted line to differentiate two sub-mechanisms investigated in their model.

The diagrams discussed in this section all serve as locality aids in constructing computational models, but then serve additional roles in determining which variables to manipulate in various simulations and in relating simulations back to the hypothesized mechanism.

Diagrams of Mechanisms to be Synthesized

Traditionally, biologists have been limited to analyzing extant mechanisms to determine what parts, operations, and organization are responsible for a phenomenon of interest. But the development of techniques for inserting genes into host organisms (typically, *E. coli*) has generated a new field of synthetic biology, in which researchers use computational models to help design regulatory networks, insert them into organisms, and assess the effects on behavior. As Cookson, Tsimring, and Hasty (2009) make explicit, diagrams play a central role in this research. In the first step "genetic wiring diagrams are translated into equations that can be analyzed." After such analysis, "modern recombinant DNA techniques are used to construct gene-regulatory networks in living cells according to the design specification." In this endeavor, diagrams are not only locality aids for developing mathematical models, but also blueprints for constructing an organism. Once the behavior of the synthesized organism can

1842

be assessed, diagrams play a further role in analyzing that behavior and revising the network design in light of the effects discovered in the synthesized organism.

This practice is illustrated in the efforts of Stricker, Cookson, Bennett, Mather, Tsimring, & Hasty (2008). They explicitly drew upon the mechanism understood to be operative in the fruit fly circadian clock to construct a synthetic clock in *E. coli*. Specifically, they added a *lacZYA* promoter to the naturally occurring *araBAD* promoter and then situated the hybrid promoter on the *araC, lacI*, and *yemGFP* genes (the last generates a green fluorescent protein used as a reporter of oscillations). Before inserting this mechanism into the bacterium, Stricker et al. constructed a diagram (Figure 5) from which they developed a computational model. Satisfied that the proposed mechanism would generate sustained oscillations under a limited set of parameter values (especially, of IPTG levels), they then employed the diagram as a blueprint for synthesizing the mechanism and as a guide to what components would have to be fine-tuned to generate sustained oscillations.
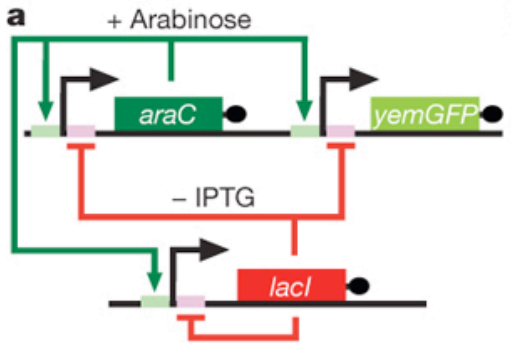

Figure 5. Stricker et al.'s (2008) diagram, which they used both to develop a computational model and to synthesize a bacterium that could generate oscillations.

The organism Stricker et al. synthesized did not behave as the model had led them to expect. Most surprising, it generated sustained oscillations under almost all parameter values tested. This led Stricker et al. to return to the mechanism as represented in the diagram and question whether processes that they had not represented in the diagram or in the equations of the model, such as protein folding, multimerization, and DNA-binding, were important to the process. They constructed a new diagram (Figure 6) and computational model that incorporated additional operations. The behavior of this model now corresponded closely to that of the synthesized bacterium. Stricker et al. concluded that the delays introduced into the feedback by these additonal steps were responsible for the oscillations.

In this example from synthetic biology, the diagram serves not only as blueprint for building the mechanism but also as a guide to determining why the mechanism did not behave as expected and then for proposing an alternative account of the mechanism.
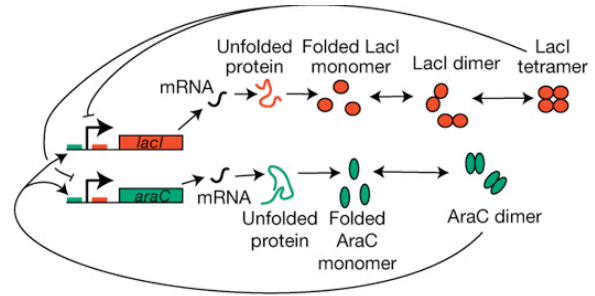

Figure 6. Stricker et al. (2008) revised diagram motivated by the discrepencies between the behavior of the synthesized organism and their computational model.

Conclusion

We have focused on one of the contexts in which diagrams provide the basis for reasoning in the development of mechanistic explanations—recomposing mechanisms through computational models and synthesized organisms. Through examples we have identified a widespread practice of constructing a diagram of the hypothesized mechanism that includes variables and parameters and using it as a locality aid in constructing equations to model the dynamics of the mechanism. But this is only the start. One of the interests in constructing a computational model is to experiment on it to determine whether the mechanism could explain various identified phenomena. A diagram can help with this, by guiding the selection of parameters to be reset or of variables to be given fixed values. When researchers set out to synthesize organisms, diagrams function both as locality aids in developing the computational models and as blueprints guiding the determination of components to include. When a synthesized organism fails to behave as the computational model suggested, researchers returned to the diagram to explore alternatives.

Our examination of published diagrams is only a first step in understanding researchers' cognitive engagement with diagrams as they seek to recompose mechanisms. Although unlikely, we cannot rule out the possibility that the diagrams we have discussed are epiphenomenal—constructed after developing the computational model as a means of communicating it to others. Given the utility of the diagram for grounding the modeling and the experiments on the model, it seems most likely the scientists would have so used it. Having identified ways diagrams appear to function in recomposing mechanisms, our hope is that other cognitive scientists will contribute to further understanding this aspect of scientific reasoning. One strategy would be ethnographic studies of modelers in which one can observe interactions with the diagrams in the process of developing and experimenting with computational models. Another strategy would involve experiments in which some modelers were allowed to create or consult diagrams while constructing a computational model and others were restricted from doing so. Such studies may help elucidate the cognitive operations that go into the construction of computational models. Further, such studies can also go beyond what we have been

able to do and address the specific features of diagrams that serve the aims of developing computational models and whether different representations, including different diagram formats, might serve these ends better. What we hope to have done is demonstrate a widespread practice of using diagrams in constructing and experimenting with computational models of biological mechanisms.

Acknowledgments

References

Bechtel, W. (2006). *Discovering cell mechanisms: The creation of modern cell biology*. Cambridge: Cambridge University Press.

Bechtel, W., & Abrahamsen, A. (2005). Explanation: A mechanist alternative. *Studies in History and Philosophy of Biological and Biomedical Sciences, 36*, 421-441.

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. (2012). Diagramming phenomena for mechanistic explanation. *Proceedings of the 34th Annual Conference of the Cognitive Science Society* (pp. 102-107). Austin, TX: Cognitive Science Society.

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.

Bünning, E. (1960). Opening address: Biological clocks. *Cold Spring Harbor Symposia on Quantitative Biology, 25*, 1-9.

Cheng, P. C.-H. (2002). Electrifying diagrams for learning: principles for complex representational systems. *Cognitive Science, 26*, 685-736.

Cheng, P. C.-H. (2011). Probably good diagrams for learning: Representational epistemic recodification of probability theory. *Topics in Cognitive Science, 3*, 475-498.

Cookson, N. A., Tsimring, L. S., & Hasty, J. (2009). The pedestrian watchmaker: Genetic clocks from engineered oscillators. *FEBS Letters, 583*, 3931-3937.

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.

Gooding, D. C. (2010). Visualizing Scientific Inference. *Topics in Cognitive Science, 2*, 15-35.

Goodwin, B. C. (1963). *Temporal organization in cells; a dynamic theory of cellular control processes*. London: Academic.

Griffith, J. S. (1968). Mathematics of cellular control processes I. Negative feedback to one gene. *Journal of Theoretical Biology, 20*, 202-208.

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.

Hegarty, M. (2004). Mechanical reasoning by mental simulation. *Trends in Cognitive Science, 8*, 280-285.

Hegarty, M. (2011). The cognitive science of visual-spatial displays: Implications for design. *Topics in Cognitive Science, 3*, 446-474.

Jacob, F., & Monod, J. (1961). Genetic regulatory systems in the synthesis of proteins. *Journal of Molecular Biology, 3*, 318-356.

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

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

Leloup, J.-C., & Goldbeter, A. (2003). Toward a detailed computational model for the mammalian circadian clock. *Proceedings of the National Academy of Sciences, 100*, 7051-7056.

Leloup, J.-C., & Goldbeter, A. (2004). Modeling the mammalian circadian clock: Sensitivity analysis and multiplicity of oscillatory mechanisms. *Journal of Theoretical Biology, 230*, 541-562.

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

Nersessian, N. (2008). *Creating scientific concepts*. Cambridge, MA: MIT Press.

Perini, L. (2005). Explanation in two dimensions: Diagrams and biological explanation. *Biology and Philosophy, 20*, 257-269.

Relógio, A., Westermark, P. O., Wallach, T., Schellenberg, K., Kramer, A., & Herzel, H. (2011). Tuning the mammalian circadian clock: Robust synergy of two loops. *PLoS Computational Biology, 7*, e1002309.

Sheredos, B., Burnston, D., Abrahamsen, A., & Bechtel, W. (in press). Why do biologists use so many diagrams. *Philosophy of Science*.

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.

Thagard, P. (2003). Pathways to biomedical discovery. *Philosophy of Science, 70*, 235-254.

Tversky, B. (2011). Visualizing thought. *Topics in Cognitive Science, 3*, 499-535.

Vitaterna, M. H., King, D. P., Chang, A.-M., Kornhauser, J. M., Lowrey, P. L., McDonald, J. D., et al. (1994). Mutagenesis and mapping of a mouse gene, *Clock*, essential for circadian behavior. *Science, 264*, 719-725.