Diagrams as Vehicles for Scientific Reasoning

William Bechtel
Department of Philosophy, Center for Chronobiology,
and Interdisciplinary Program in Cognitive Science
University of California, San Diego

Adele Abrahamsen
Center for Research in Language
University of California, San Diego

Abstract

We argue that diagrams are not just a communicative tool but play important roles in the reasoning of biologists: in characterizing the phenomenon to be explained, identifying explanatory relations, and developing an account of the responsible mechanism. In the first two tasks diagrams facilitate applying visual processing to the detection of patterns that constitute phenomena or explanatory relations. Diagrams of a mechanism serve to guide reasoning about what parts and operations are needed and how potential parts of the mechanism are related to each other. Further they guide the development of computational models used to determine how the mechanism will behave. We illustrate each of these uses of diagrams with examples from research on circadian rhythms.

Key Words: Diagrams; visual reasoning; mechanistic explanation; phenomena; explanatory relations; circadian rhythms

Introduction

Anyone who has read a textbook or journal article or attended a talk by a biologist knows that biologists make extensive use of diagrams. What function do these serve? One possibility is that their sole function is communication, in that scientists deploy them late in the research process as a tool for conveying hypotheses, findings, and other completed aspects of the research process to an audience. If so, those interested in the cognitive activities of scientific inquiry may regard them as epiphenomenal. We will argue that in fact diagrams serve a variety of additional functions, and will focus on how they are employed in the central explanatory activities of scientists: characterizing the phenomenon to be explained, identifying explanatory relations, and developing an account of the responsible mechanism.[1]

---

[1] Since Larkin and Simon (1987) pioneering research on how diagrams convey information differently than textual representations, a number of cognitive scientists have contributed to our understanding of the various devices used in diagrams to convey information (Tversky, 2011) and how people reason with diagrams (Hegarty & Just, 1993; Hegarty, 2011). Others have focused on how diagrams can be better designed to aid learning in fields such as electricity and probability theory (Cheng, 2002, 2011). Most closely related to

A major goal in many fields of biology is to explain a phenomenon by characterizing the mechanism taken to be responsible for generating it. Mechanistic explanation involves decomposing the posited mechanism into its parts and the operations they perform, and recomposing it so as to capture how the parts are organized and the operations coordinated so as to generate the phenomenon.[2] Although there are many differences between machines designed by human designers and mechanisms proposed by scientists, they share the idea that the coordinated execution of operations performed by parts is what is responsible for the generation of the phenomenon. Just as machine designers often use diagrams to explore and reason about possible designs, biologists often use what we call *mechanism diagrams*. These are crucial tools for advancing possible mechanisms, guiding experimental research or computational modeling needed to assess and refine the proposed mechanisms, enabling exploration of alternative mechanisms, and finally presenting the (currently) accepted explanation. We will explore how mechanism diagrams figure in several of these various activities.

Before turning to that, we will explore two other uses of diagrams that are equally important to scientists as they develop mechanistic explanations. The first use is in delineating the phenomenon to be explained. Mechanisms are typically identified in terms of the phenomenon for which they are taken to be responsible. This is important since typically in living organisms mechanisms are not discretely packaged but rather rely on parts that are distributed throughout the organism – parts that also may figure in generating other phenomena as parts of different mechanisms. It is the scientist who selects out of the multitude of interacting components of an organism those that can be identified as contributing to a given phenomenon and thus taken to be working together as the responsible mechanism. Phenomena, too, are not discretely packaged. They are identified by scientists who pick out and treat as discrete phenomena patterns of activity that they view as sufficiently coherent and important to merit explanatory effort. As we will see, they often use what we call *phenomenon diagrams* to identify these patterns.

---

our objective in this paper are those who have investigated the roles of specific diagrams in the reasoning of scientists such as Galileo (Cheng & Simon, 1995) and Maxwell (Nersessian, 2008). Gooding (2004, 2010) in particular focuses on how scientists develop representational formats that enable them to our visual processing capacities to elicit patterns in tasks such as understanding extinct organisms. What distinguishes our analysis is that we focus on the roles diagrams play in developing mechanistic explanations that figure prominently in biology.

[2] While mechanistic explanations were pursued by biologists throughout the 19th and 20th centuries, they were little discussed in the dominant approaches to philosophy of science in the 20th century, which emphasized derivation from laws as the main explanatory activity. Recognizing that deriving results from laws plays little role in biology, recently a number of philosophers have attempted to characterize mechanistic explanations as pursued by biologists, including Bechtel and Richardson (1993/2010), Bechtel and Abrahamsen (2005), Craver (2007), and Machamer, Darden, and Craver (2000).

Often phenomena are characterized in terms of relations between variables and identifying phenomena is a matter of highlighting specific relations between variables. Even after conceptualizing a phenomenon, researchers continue to focus their research on identifying relations between variables, but now focusing on those variables that are taken to correspond to properties of the parts and operations in the mechanism responsible for the phenomenon. This may take the form of relating one or more variables employed in characterizing the phenomenon to ones thought to describe other entities that could be part of the responsible mechanism. It may also involve showing that putative explanatory variables exhibit the same pattern as found in the phenomenon itself (e.g., the explanatory variables oscillate in the same way as the variables characterizing the phenomenon). Scientists have developed a host of diagramming techniques to depict these relations. The resultant *explanatory relations diagrams* not only figure in the initial development of mechanism diagrams but continue to play fundamental roles after mechanisms have been proposed and even agreed to by the scientific community as often it is the specific relations that are central to the inquiry at hand.

Our objective in discussing these three types of diagrams is to show that they function not solely to communicate to others but also in the reasoning processes of scientists who produce and use them. The relevant reasoning processes prominently include those of visual processing, which are particularly effective in picking out patterns that are much harder to discern when data are presented in text or tables. Researchers exploit this characteristic of visual reasoning by designing representational formats that are well suited for picking out the kinds of patterns that are relevant in their field.

Throughout this paper we will use the research field of circadian rhythms as an exemplar. As the name (*circa* = about + *dies* = day) suggests, circadian rhythms are oscillations on an approximately 24-hour cycle. They are generated endogenously within organisms, yet entrainable to the day-night cycle in their local environment. These oscillations have been studied in organisms ranging from cyanobacteria and fungi to plants and animals. While in humans they are perhaps most widely associated with sleep patterns, they regulate a broad range of physiological and behavioral activities, generally by regulating the expression of particular genes. To facilitate discussion without having to introduce too much biological detail, we will focus on the mechanisms involved in the two most-studied animal systems: fruit flies and mice. These mechanisms are thought to have origins in a common ancestor and so contain many homologous components that are referred to by the same name. We will take note when we shift from flies to mice, but not emphasize the differences.

Phenomenon Diagrams

In presenting phenomena as the targets of explanation, Bogen and Woodward (1988) distinguished phenomena from the data they generate. On their account, a phenomenon is a repeatable regularity, and data are observed instances that point to or provide evidence for the phenomenon. Although Bogen and Woodward treat phenomena "as in the world, as belonging to the natural order itself and not just to the way we talk about or conceptualize that order" (p. 321), it is important to note that it is only through cognitive activity that researchers arrive at the patterns or regularities that they regard as phenomena to explain.

Many of these activities take advantage of graphing or diagramming formats that display data in ways that lead our visual system to pick up patterns. The scientist may posit a phenomenon when similar patterns are found across multiple occasions. When a mechanism is proposed to explain the phenomenon, an important task in evaluating the proposed mechanism is to demonstrate, either qualitatively or quantitatively, that it is capable of producing the general pattern. Turning to circadian rhythms we can see how diagrams have served to represent and characterize specific circadian phenomena.

Observations of circadian rhythms have been reported since ancient times when people noticed daily cycles in specific behaviors of organisms, but detailed recordings yielding data in which patterns could be identified began only in the 18th and 19th centuries. De Mairan (1729) recorded the daily opening and folding of the leaves of Mimosa plants to determine whether this behavior was generated internally or was a response dependent on exposure to light. He placed the plants in a dark cupboard and showed that they still opened and closed their leaves daily. These observations did not require quantitative analysis, but demonstration of other circadian phenomena did. To show that there is an underlying 24 hour oscillation in body temperature to which random perturbations are appended, Wunderlich (1868) collected multiple recordings per day from over 25,000 individuals. His results not only established 37° C (98.6° F) as normal mean body temperature, but also revealed daily oscillations of over 1° C between a low in the early morning and a peak in the afternoon.

Wunderlich presented his results only in summary tables, but it requires effort to discern an oscillatory pattern in tabular data. Line graphs are more commonly employed to make the pattern of oscillation manifest to our visual system. In Figure 1, for example, each data point (small square) indicates the body temperature of one person as measured at 20 different times across two nights and days. Connecting data points with lines (including dashed lines during periods of sleep, when no recordings were made) results in a figure in which one can readily perceive the oscillatory pattern in body temperature. Superimposing light and dark bars on the abscissa's timescale and shading the corresponding areas of the graph visually highlight the pattern of interest: body temperature drops towards and during the night and increases during the day. Such graphs can be extended to show multiple individuals (or the same individual during different epochs), rendering it relatively easy to distinguish regular patterns or unusual instances. The regular patterns are taken as phenomena to be explained.

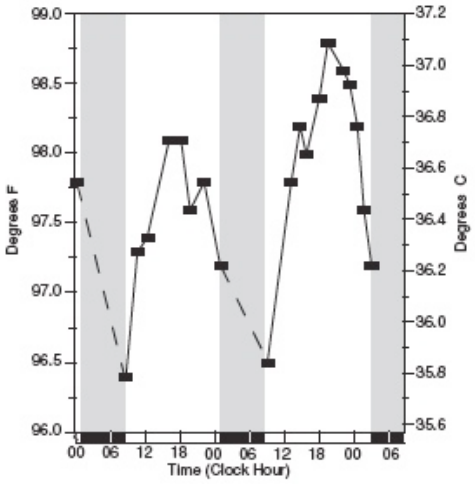

Figure 1. Life graph from Koukkari and Southern (2006) exhibiting the circadian oscillation in body temperature of an individual person.

Although this line graph makes clear that there is a circadian pattern of rising and falling body temperature, it does not render transparent related phenomena that circadian researchers regard as critical, such as the exact duration of each cycle of the oscillation (its period) and whether the timing of the onset of each cycle remains constant across multiple days Accordingly, researchers have developed a number of other formats for displaying data. A particularly important type of diagram for circadian rhythm research is the *actogram*, in which short vertical marks represent each act or each above-threshold value on some measurement, going left to right across the day-night cycle ("day"). The data for successive days are stacked from top to bottom, making it easy to spot stability, trends, or disorder. (Where the vertical marks are dense, some investigators simplify the plot by substituting a solid horizontal bar.)

Figure 2 shows a typical contemporary actogram that displays 50+ days of a mouse's locomotor behavior. The first few days had a normal alternation of Light-Dark (LD), as specified by the light and dark portions of the bar at the top. On the remaining days the mouse was in constant darkness (DD). The arrow on the right (LP) indicates a day on which a light pulse was administered. At the bottom actual clock time is shown. This actogram is *double plotted*: each line shows 48 hours of recording, with the second 24-hour period repeated at the beginning of the following line. This convention developed to make it easy to detect patterns even when the activity phase would extend over the edge of the actogram if only 24 hours had been shown. Visual inspection makes clear that when this nocturnal animal is experiencing a light-dark regime, its activity occurs nearly exclusively during the dark phase. Once in constant darkness (a condition known as free-running), the animal begins its activity somewhat earlier each day. The light pulse produces a delay in activity on the subsequent day, but thereafter the pattern of beginning activity earlier each day resumes. The actogram makes circadian behavior, including periods of less than or greater than 24 hours, immediately detectable by visual inspection.

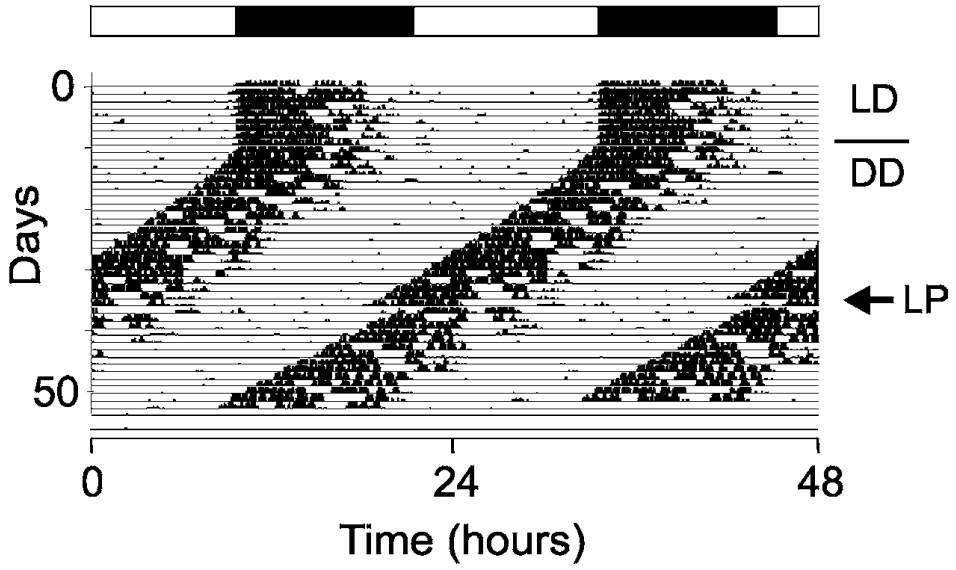

Figure 2. Example from Lowrey & Takahashi (2004) of a contemporary actogram. It makes apparent how the activity of a mouse is entrained to light when it is under a light-dark cycle during the initial days and free runs with a period somewhat less than 24 hours when in constant darkness.

Representational vehicles such as the actogram often undergo refinement over time. By examining their history we can see which ways of presenting information turned out to be useful and which were discarded. The first actogram we have been able to identify was produced by Johnson (1926), who was investigating the nocturnal versus diurnal behavior of various mammals. Johnson devised the use of a disk rotated by a clock on which movements of a mouse in a cage were recorded as deflections in an otherwise smooth tracing (Figure 3, left side). While one could compare multiple circular tracings to assess changes or stability over successive days, Johnson found it more informative to unroll each circular tracing (one day's data) into a straight line and place each tracing below the previous one such that the hours of all days were in alignment (Figure 3, right side). Along the right edge of the actogram Johnson labeled the various conditions the animal experienced. Recording began when the animal was exposed to constant daylight. The condition then shifted to constant darkness, then to a reversed light-dark cycle (light at night), and finally to constant darkness again. It can be seen that the animal's normal oscillatory pattern of behavior, established prior to the investigation, essentially continued in total light or total darkness but underwent a fairly rapid reversal when exposed to light at night.

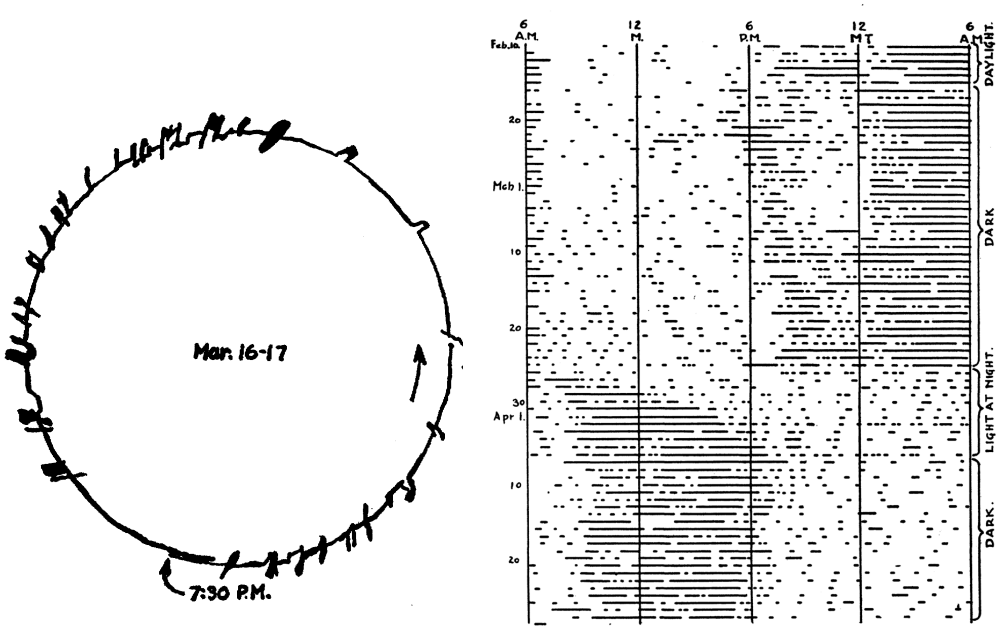

Figure 3. The earliest example of an actogram from Johnson (1926). On the left a rotating disk records any time the animal moved; each day's tracing is unrolled to generate a line on the actogram on the right.

There are several ways in which modern actograms, such as that in Figure 2, render more clearly the behavioral patterns that require explanation. First, the use of a light-dark bar makes it easier to relate the activity to the current or recent pattern of light exposure. Second, double plotting makes it easier to see the beginning and end of behavioral bouts that straddle the 24th hour. Third, many actograms follow the practice illustrated in Figure 1 of using shading to make it even more visually salient at what times the animal is exposed to darkness.

While the changes in the actogram from its first inception to current examples are illustrative of the type of steps researchers take in the quest to render phenomena visually apparent, another vehicle for representing circadian phenomena, the *phase response curve*, illustrates this even more powerfully. When investigating how circadian rhythms can be entrained by cues such as light, researchers frequently keep animals in darkness but expose them to pulses of light at particular times. They then record how much the onset phase of their rhythm is advanced or delayed by light pulses at selected intervals and display these discrepancies in a phase response curve. (The actogram can show the response to specific light pulses, but not the pattern of how organisms change in response to light pulses at different times of day.)

In a pioneering study of such entrainment effects, Hastings and Sweeney (1958) worked with *Gonyaulax polyedra*, a photosynthetic marine dinoflagellate that produces luminescence when disturbed. They grew these organisms first under a normal light-dark cycle and then in constant darkness. Next they exposed them to a three-hour pulse of light, varying the onset time of the pulse so as to determine how much that shifted the time of maximal luminescence. They presented their results in the diagram on the left in Figure 4.

Hour 0 is the onset of constant darkness. Vertical lines at 7, 31, and 55 hours indicate the time of maximal luminescence in control organisms (who were not exposed to the pulse). The horizontal lines represent organisms exposed to the three-hour pulse at different hours of delay after the onset of darkness (3, 7, 11, . . .). The researchers fitted curves to data points marked by tiny triangles, which indicate the time of the organism's subsequent maximum luminescence. The distance of each data point from the nearest vertical line represents the extent of advance or delay. It can be seen by following the horizontal line labeled 23 on the scale on the left that organisms exposed to a 3-hour light pulse beginning 23 hours after onset of darkness show maximum luminescence at hour 32 rather than 30, a phase delay of 2 hours. In contrast, pulses beginning 7 hours after darkness produce a large phase advance.

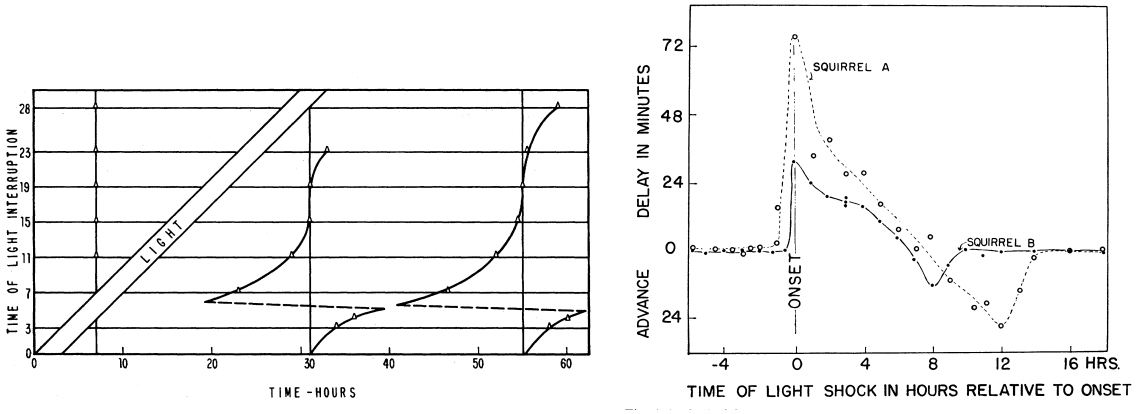

Figure 4. Left: The first phase response curve produced by Hastings and Sweeney (1958) in which phase delays and advances are shown by how far the curved lines depart from the vertical ones. Right: An alternative format for a phase response curve introduced by DeCoursey (1960) which has become standard in showing advances or delays as points below or above 0 on the y-axis.

While this diagram does present the crucial information, the pattern is not easy to recognize since one must calculate how far, at each hour, the curves differ from the vertical lines. Thus, it is not surprising that just a couple of years later DeCoursey (1960) introduced an alternative phase response curve that uses a different format to represent the same information (right side of Figure 4). In her study of flying squirrels kept in constant darkness, she indicated on the abscissa the time of a ten-minute light pulse relative to the usual onset time of their running-wheel activity. The data points indicate the consequent advance (deflection below 0) or delay (deflection above 0) in onset of running for two squirrels. With this representation, it is easy to see that light around the beginning of this nocturnal animal's usual activity period (the onset of night) delays its activity, whereas light 8 to 12 hours later (the end of night) advances its activity. Light during its daytime hours (from approximately 12 hours to 0 hours) has no effect. Having represented the phenomenon this way, one can also readily make sense of it—daytime light exposure does not indicate a need to reset the phase of one's activity, whereas light imposed in early nighttime indicates either that the endogenous rhythm is out of synchrony with the external environment or that the period of daylight has expanded. The appropriate adjustment is to delay activity. Likewise, late-night light indicates a need to stop its activity

sooner. In the decades since DeCoursey introduced this format the major change to it has been to reverse the y-axis so that advances are up and delays are down.

What line graphs, actograms, and phase response curves all provide are ways for scientists to make visually apparent the patterns in data that constitute phenomena to be explained. Each technique allows plotting data in such a manner as to make patterns constituting different aspects of circadian behavior apparent to our visual system.

Explanatory Relations Diagrams

While an important goal for many circadian rhythm researchers is to develop mechanistic explanations of particular circadian phenomena, their research papers often include little discussion of the overall mechanism. Typically just one (or none) of the figures in the paper depict the parts and operations of the hypothesized mechanism. Rather, most of the reasoning in the paper involves hypothesizing and offering evidence for relations between variables that quantify properties of the mechanism's parts and operations. Examples of such mechanism variables include the concentration of the protein PER (a property of one part of a molecular feedback mechanism discussed below) and the rate of transcription of the gene *per* (a property of an operation of another part of the same mechanism). For a given mechanism, some of its variables exhibit systematic relations with each other. The goal of studying and presenting these relations is to uncover how the relevant parts or operations depend on each other, which contributes to the greater goal of achieving a full mechanistic explanation of the phenomenon of interest. Note that the phenomenon itself often is expressed as a relation between variables, but in that case the variables are ones that capture aspects of the overall activity of the mechanism and/or its environment. One example, described above, is the effect of the time of a light pulse on the number of minutes of delay (or advance) in an organism's activity of running. Finally, mixed relations are sometimes of interest, such as finding similarly timed oscillations in an organism's activity and in PER concentrations within its cells. We refer to diagrams of relations that include at least one mechanism variable as *explanatory relations diagrams* , because such relations are of interest due to their potential to help explain a phenomenon.

In one of the first studies to investigate the role of genes in generating circadian rhythms, Konopka and Benzer (1971) examined the relation between various mutant strains of fruit flies and two measures of circadian rhythmicity. He first focused on the number of flies that eclode (emerge from their pupa) at different times of day. As shown in the top graph on the left in Figure 5, under free-running conditions (constant darkness, temperature, etc.) most flies eclode in the early part of the day. The lack of light cues presumably explains the spread of the period of high eclosion in later days. Konopka and Benzer compared three mutant strains they generated with the flies exhibiting normal circadian rhythms. It is easy to see from looking at the line graphs in panels B-D and comparing them with panel A that the different strains exhibited loss of rhythm, a shortened period (~19 hours), or a long period (~28 hour).

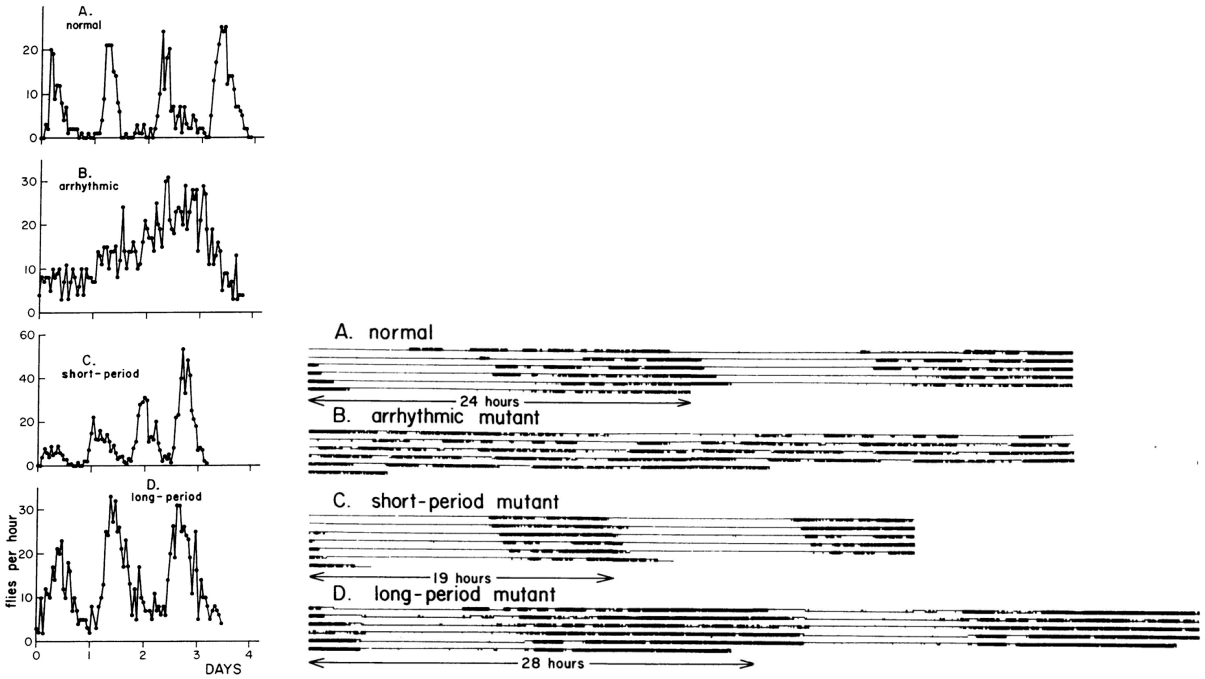

Figure 5. Left: Line graph from Konopka & Benzer (1971) showing circadian rhythms of eclosion from a population of normal flies (A) and three mutant populations (B-D). Right: actogram showing the activity periods of normal flies (A) and three mutant populations (B-D).

Since eclosion is a once-in-a-lifetime activity, Konopka and Benzer also examined the locomotor activity of the flies under free-running conditions (locomotor activities was detected when flies disrupted beams of infrared light that does not affect their eyes). This also served to demonstrate that the mutation affected central timekeeping, not just the specific mechanism of eclosion. They reported these results in a set of actograms that showed the behavior of one fly exhibiting normal circadian rhythms and one fly exhibiting each mutation. The actograms make it visually apparent that the mutant flies lack rhythms or have short or long periods. By changing the period recorded on each line to 38 hours (two successive 19 hour periods) in panel C and to 56 hours (two successive 28 hour periods) in panel D, the authors show that the mutant flies maintain regular shortened or lengthened period rhythms. This has the effect, however, of making less apparent that the flies are beginning their activities much earlier or much later each day than if they had used the same 48 period used in panels A and B. This provides a reminder that a strategy that is effective for showing one pattern may make another less obvious.

Konopka and Benzer labeled the gene that they mutated *period*, but until the development of cloning techniques in the 1980s it was not possible to identify or measure concentrations of either the mRNA or protein produced by this gene. When cloning became available, researchers began to measure concentrations of both mRNA transcripts and the resulting protein. In a pivotal study, Hardin, Hall, and Rosbash (1990) set out to relate concentrations of *per* mRNA to whether the fly exhibited normal rhythms or was a short or long period mutant. (They assessed the amount of *per* mRNA by applying the RNase

protection procedure to total RNA extracted from the heads of flies selected to be sacrificed each hour.) They presented their results in both Western blots and in line graphs as shown in Figure 6. On the left are results when the flies experienced 12 hours of light followed by 12 hours of darkness (LD) and on the right are the results for continuous darkness (DD). For the normal flies (*per*$^{+}$) in LD, RNA levels rose late in the day, peaked as lights were turned out (at 12 H circadian time, which is measured from the beginning of the light cycle) and remained high for 5-8 hours before falling to their lowest levels 1-2 hours before lights were turned back on. The peak for the short period mutant (*per*^{S}) is earlier than that of the normal flies while that for the long-period mutant (*per*^{L}) is delayed. In the right panel they show that RNA oscillations under continuous darkness have an approximate period of 24 hours in normal flies but only 20 hours in short-period mutants.

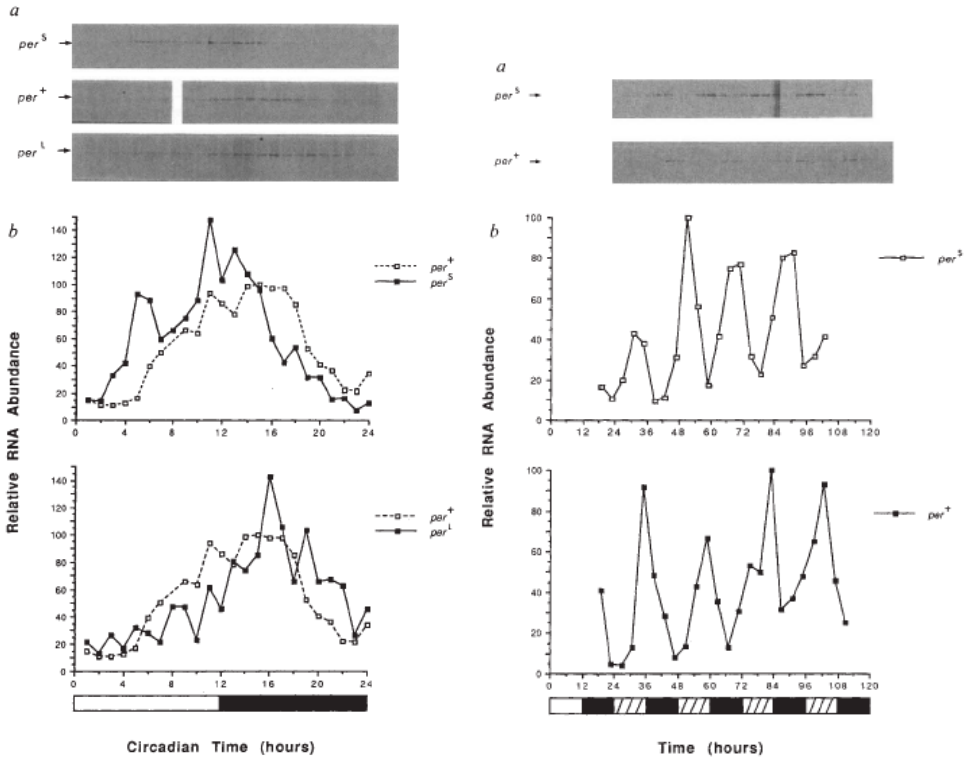

Figure 6. Western blots and line graphs of normal flies (*per*$^{+}$) and two mutant populations (*per*^{S} and *per*^{L}) showing RNA concentrations over 24 hours (Left) and four days (Right).

None of the diagrams in the Hardin et al. paper relate oscillation of *per* mRNA to features of circadian phenomena such as the timing of eclosion or activity periods. Instead Hardin et al. note that these results are very similar to those that Konopka and Benzer had generated for behavior. They clearly view the pattern of expression of *per* mRNA in the different fly strains as explanatory of the circadian rhythms detected in behavioral measures without actually showing that in a diagram.

The diagrams we have exhibited in this section are clearly intended to provide explanation. Konopka and Benzer's goal in identifying a gene in which mutations can result in aberrant

circadian rhythmicity was to identify a potential part of the mechanism responsible for producing circadian rhythms in the fly's activity—either the gene itself or some product of the gene. By demonstrating that *per* mRNA (and PER protein) concentrations oscillate and that these molecular oscillations show altered timing in mutants that is similar to the altered timing of their behavioral rhythms, Hardin et al. identified *per* transcription and translation as operations in the molecular mechanism underlying those rhythms. What we have emphasized is how they use diagrams to render the proposed explanatory relation visually apparent. Thus, Konopka and Benzer showed both in line graphs and actograms how mutations altered the phenomena of interest. Hardin et al. used both blots and line graphs to show that the same altered pattern found in the behavior of the mutants could be found in the levels of mRNA. As circadian researchers have followed up on these initial studies in the past 20 years they have advanced many more explanatory relations, which they typically make visually perspicuous by using diagrams. Moreover, in the process they have developed many additional diagram formats for displaying explanatory relations. For example, to determine whether different variables are synchronized they use not only line graphs, but also Rayleigh plots and raster plots. (We cannot pursue those examples here, but see the figures in Maywood, Reddy, Wong, O'Neill, O'Brien, McMahon, Harmar, Okamura, & Hastings, 2006).

Mechanism Diagrams

The explanatory relations discussed in the previous section are commonly viewed as indicating causal relations (possibly very indirect) between the entities whose properties correspond to the variables shown to be related. But biologists typically do not regard individual causal interactions as providing the full explanation for the phenomenon they are seeking. Rather, they view these interactions as occurring between parts or operations within a mechanism or between these components of the mechanism and the phenomenon of interest. The ultimate goal for many biologists is to identify the parts, operations, and organization of the mechanism and understand how it generates the phenomenon. Thus, as they investigate explanatory relations, they have in mind a conception of the mechanism or, on the basis of their research, put forward a proposal as to the mechanism. The conception of the mechanism is often developed in diagrams in which icons represent various parts and arrows and other devices represent operations through which parts are transformed or affect other parts. Although limited to the two dimensions of the page, researchers use spatial relations sometimes to indicate actual spatial locations of parts and operations but also to cluster parts and operations that affect each other so as to form a conceptual perspective on how the mechanism works. Researchers turn to these spatial representations both to put explanatory relations into context and to reason about what additional relations need to be investigated.

Mechanism diagrams appear relatively infrequently in published research. While most diagrams in a paper serve to identify the explanatory relations that the research has identified, a mechanism diagram typically only appears as a first or last figure. As the first figure it often serves to present the received conception of a mechanism that situates the experiments to be performed in the study. As a final figure it represents the conception of the mechanism that has been arrived at on the basis of the research. If the paper is

accompanied by an introductory paper in the same issue of the journal, as sometimes is the case in journals such as *Science* or *PNAS*, the mechanism diagram typically will appear in that paper, not the research paper. Mechanism diagrams also often occur in review articles, but even there they tend to be outnumbered by diagrams demonstrating explanatory relations. This may suggest that mechanism diagrams are merely communication devices that don't play an important role in the reasoning of scientists. Scientists themselves, however, frequently draw mechanism diagrams as they are developing and evaluating hypotheses. These diagrams serve as a basis for exploring possible configurations of a mechanism and developing further empirical inquiry (Burnston, 2013), as the framework in which different explanatory relations are integrated, and as the basis for proposing additional components to the mechanism or relations between one mechanism and others.

Building on their finding of 24-hour oscillations in the concentrations of *per* mRNA and PER protein in the fruit fly, Hardin et al. proposed a mechanism whereby the operations of transcribing and translating *per* could generate such oscillations. The key idea was that in some way the protein PER figured in a feedback process whereby it inhibited its own further production. It was known that a feedback loop was the type of organization that could generate oscillations, and that with sufficient delays and non-linearities these oscillations could be sustained. Hardin et al. presented their conception of a feedback mechanism in Figure 7.

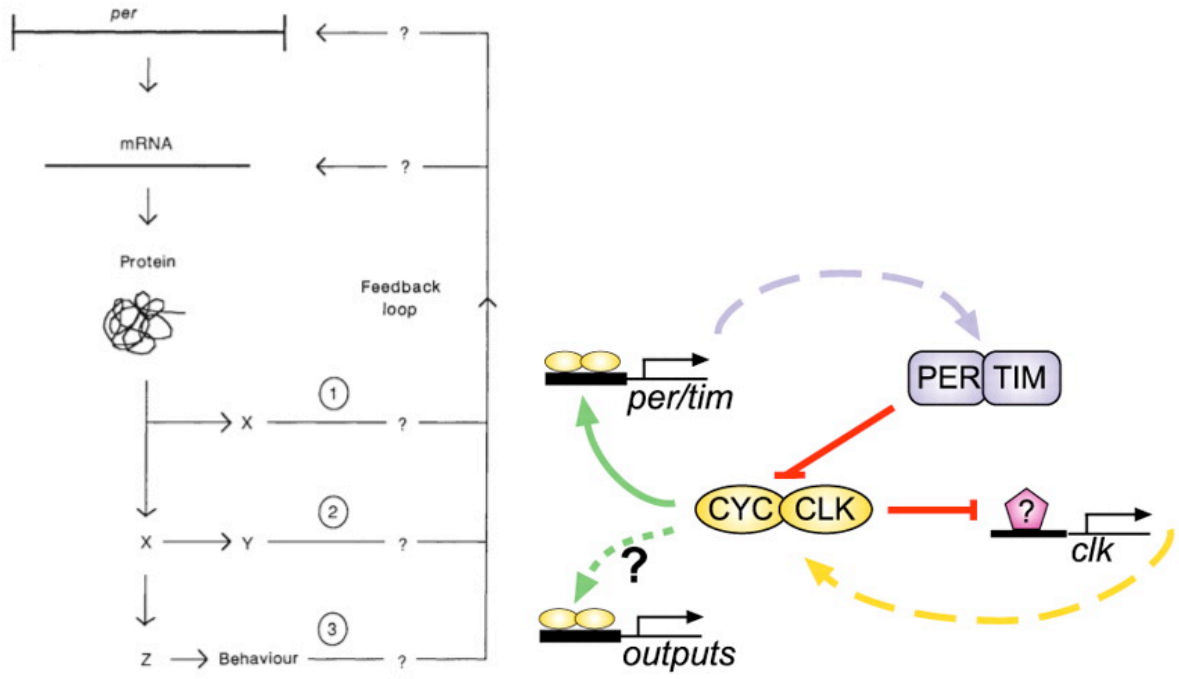

Figure 7. Left: Hardin, Hall, and Rosbash's (1990) mechanism diagram proposing a transcription-translation feedback loop for generating circadian rhythms in fruit flies. Right: Harmer, Panda, and Kay's (2001) diagram showing how the understanding of the fruit fly clock mechanism had developed over the following decade. Note the prominent use of question marks in both figures.

A notable feature of this and many mechanism diagrams in biology is the inclusion of question marks. These are a strong signal that the researchers are using the diagram as an aid for reasoning about a mechanistic explanation and that it is still in flux. Hardin et al. were committed to some sort of negative feedback mechanism, but recognized that they could not yet determine what was being fed backwards and where it affected other processes. Thus, they proposed three possible origins of the feedback—the protein itself, another chemical compound generated by the protein, or some behavior of the organism that relied upon the protein. They also specified two possible targets of the negative feedback, either of which would inhibit protein production: the transcription of *per* or its translation into PER protein. Over succeeding years, researchers determined that in fact PER was transported back into the nucleus where it could inhibit transcription. It was further established that PER itself was transported back into the nucleus as part of a dimer with another protein, Timeless (TIM), and that neither PER nor TIM had a DNA-binding region. Instead they were found to interact with another dimer consisting of CYCLE (CYC) and CLOCK (CLK), a more complicated resolution of the upper question mark than they had anticipated. Although the concentration of CYC was shown not to oscillate, the fact that CLK exhibited circadian oscillations, suggested another feedback operation. In their more refined representation of the mechanism in 2001 (shown on the right in Figure 7), Harmer, Panda and Kay included these now-understood operations without question marks. Instead, question marks denote two remaining gaps in knowledge—what bound to the promoter of *clk* and what served as the output that enabled this mechanism to regulate expression of other genes in the organism.

One of the challenges in understanding a mechanism is figuring out how the operations proposed in the mechanism give rise to the phenomenon of interest, in this case circadian rhythms. When mechanisms have a small number of operations organized in a simple sequence, researchers can mentally simulate its activity. Looking at the upper left portion of Figure 7, one can envisage the gene being transcribed and translated so that over time the concentration of the protein increases. As it does so, one envisages the feedback process starting to have an effect in reducing the rate of transcription. Since chemical compounds gradually degrade (not shown in the diagram), one can ascertain that the levels of PER will decrease, allowing the rate of transcription to increase again. As useful as a mental simulation of this sort may be, it at best provides a qualitative understanding. From the mechanism diagram and mental simulation alone one cannot evaluate whether the oscillations will be regular and sustained long enough to count as circadian rhythms. The strategy to which scientists turn is to represent the mechanism mathematically is computational modeling.

In order to develop a computational model grounded in an understanding of the mechanism, researchers must construct equations in terms of variables and parameters to describe how substances that change in quantity do so. Modelers often invoke mechanism diagrams in this effort, where these diagrams function 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). To model the mechanism proposed by Hardin et al., Goldbeter (1995) constructed the diagram shown on the left in Figure 8 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 through empirical inquiries to be necessary before PER could be transported back into the nucleus). Relying on the diagram as a locality aid, Goldbeter constructed a set of differential equations, each characterizing the change in concentration of one of the molecular components. On the basis of generating sustained oscillations by running the model with appropriate parameter values, Goldbeter argued that the proposed mechanism could generate sustained circadian oscillations. As more molecular components of the mechanism were discovered during the 1990s (resulting in conceptions such as exhibited in the Harmer et al. diagram on the right in Figure 7), Goldbeter built more elaborate computational models in much the same manner, employing a diagram to identify variables and parameters. He used these models to show not just that the proposed mechanism could generate circadian rhythms, but how it could be entrained by light and, when appropriately altered, could exhibit the patterns of known circadian pathologies such as advanced sleep phase syndrome (Leloup & Goldbeter, 2000; Leloup & Goldbeter, 2003, 2004).

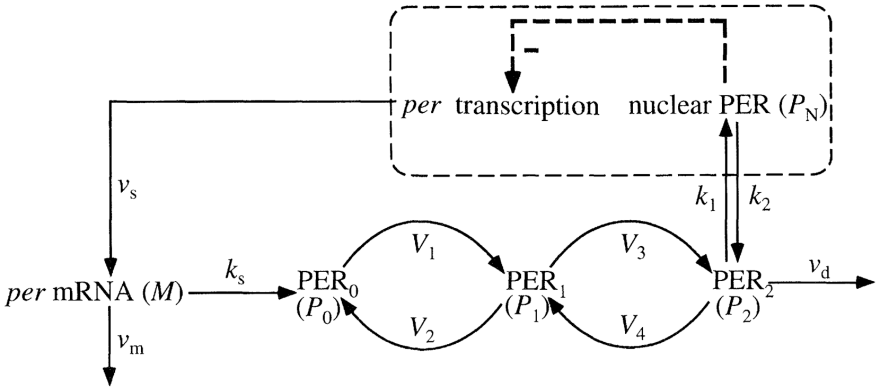

Figure 8. Mechanism diagram Goldbeter (1995) developed as a locality aid in constructing a computational model of the transcript-translation feedback loop in the fruit fly as then understood.

In this section we have introduced how diagrams contribute to the investigation of mechanisms by providing a framework in which to develop hypotheses about the mechanism and computational models that can ascertain whether it could generate the phenomenon. In both of these tasks the process of spatially arranging parts and operations in a diagram plays important functions. First, it forces researchers to think in terms of the continuity of the operations in the mechanism. If one doesn't know how to connect operations, then inserting a question mark indicates either a gap or one of the possible ways to fill a gap. Second, it provides a locality aid to organize the variables and parameters the need to be related in equations in order to simulate the mechanism's functioning.

Conclusions

We have exhibited how diagrams serve as vehicles in the reasoning of scientists by focusing on three different roles they play in the development of mechanistic explanations. First, diagrams figure prominently in delineating the phenomenon to be explained by enabling scientists to plot data in ways that render it visually apparent. Second, diagrams are used to represent explanatory relations by showing (a) how a potential component of the responsible mechanism is related to an aspect of phenomenon; or (b) how the components of the mechanism generate a pattern characteristic of the phenomenon. We emphasized that researchers often develop multiple representations to make apparent different relationships revealed by the data. Finally, diagrams are used to show the parts, operations, and organization that are proposed to explain the phenomenon. We highlighted the role question marks play in focusing on questions yet to be answered and how mechanism diagrams provide the basis for computational models designed to show that the mechanism could indeed generate the phenomenon. In all cases, diagrams serve to engage visual processing that figures in the reasoning activities of scientists.

References

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

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.

Bogen, J., & Woodward, J. (1988). Saving the phenomena. *Philosophical Review, 97*, 303-352.

Burnston, D. C. (2013). Mechanism diagrams as search organizers. *Proceedings of the 35th Annual Conference of the Cognitive Science Society* (pp. 1952-1957). Austin, TX: Cognitive Science Society.

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.

Cheng, P. C.-H., & Simon, H. A. (1995). Scientific discovery and creative reasoning with diagrams. In S. M. Smith, T. B. Ward & R. A. Finke (Eds.), *The creative cognition approach* (pp. 205-228). Cambridge, MA: MIT Press.

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

De Mairan, J.-J. d. O. (1729). Observation Botanique. *Histoire de l'Academie Royale Sciences*, 35.

DeCoursey, P. J. (1960). Daily light sensitivity rhythm in a rodent. *Science, 131*, 33-35.

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. (2004). Cognition, Construction and Culture: Visual Theories in the Sciences. *Journal of Cognition and Culture, 4*, 551-593.

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

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.

Harmer, S. L., Panda, S., & Kay, S. A. (2001). Molecular bases of circadian rhythms. *Annual Review of Cell and Developmental Biology, 17*, 215-253.
Hastings, J. W., & Sweeney, B. M. (1958). A persistent diurnal rhythm of luminescence in *Gonyaulax polyedra*. *Biological Bulletin, 115*, 440-458.
Hegarty, M. (2011). The cognitive science of visual-spatial displays: Implications for design. *Topics in Cognitive Science, 3*, 446-474.
Hegarty, M., & Just, M. A. (1993). Constructing Mental Models of Machines from Text and Diagrams. *Journal of Memory and Language, 32*, 717-742.
Johnson, M. S. (1926). Activity and distribution of certain wild mice in relation to biotic communiities. *Journal of Mammalogy, 7*, 254-277.
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.
Larkin, J. H., & Simon, H. A. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science, 11*, 65-99.
Leloup, J.-C., & Goldbeter, A. (2000). Modeling the molecular regulatory mechanism of circadian rhythms in *Drosophila*. *BioEssays, 22*, 84-93.
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.
Lowrey, P. L., & Takahashi, J. S. (2004). MAMMALIAN CIRCADIAN BIOLOGY: Elucidating Genome-Wide Levels of Temporal Organization. *Annual Review of Genomics and Human Genetics, 5*, 407-441.
Machamer, P., Darden, L., & Craver, C. F. (2000). Thinking about mechanisms. *Philosophy of Science, 67*, 1-25.
Maywood, E. S., Reddy, A. B., Wong, G. K., O'Neill, J. S., O'Brien, J. A., McMahon, D. G., Harmar, A. J., Okamura, H., & Hastings, M. H. (2006). Synchronization and maintenance of timekeeping in suprachiasmatic circadian clock cells by neuropeptidergic signaling. *Current biology : CB, 16*, 599-605.
Nersessian, N. (2008). *Creating scientific concepts*. Cambridge, MA: MIT Press.
Tversky, B. (2011). Visualizing thought. *Topics in Cognitive Science, 3*, 499-535.
Wunderlich, K. R. A. (1868). *Das Verhalten der Eigenwärme in Krankheiten*. Leipzig: Otto Wigard.