Rev.Phil.Psych. (2015) 6:117–131
DOI 10.1007/s13164-014-0215-2

# Diagrams as Tools for Scientific Reasoning

**Adele Abrahamsen · William Bechtel**

Published online: 12 November 2014


**Abstract** We contend that diagrams are tools not only for communication but also for supporting the reasoning of biologists. In the mechanistic research that is characteristic of biology, diagrams delineate the phenomenon to be explained, display explanatory relations, and show the organized parts and operations of the mechanism proposed as responsible for the phenomenon. Both phenomenon diagrams and explanatory relations diagrams, employing graphs or other formats, facilitate applying visual processing to the detection of relevant patterns. Mechanism diagrams guide reasoning about how the parts and operations work together to produce the phenomenon and what experiments need to be done to improve on the existing account. We examine how these functions are served by diagrams in circadian rhythm research.

## 1 Introduction

Anyone who has read a journal article or attended a talk by a biologist knows that biologists make extensive use of diagrams. What functions do these serve? The most obvious function is communication. Late in the research process, diagrams are deployed in research reports to convey to others the hypotheses, apparatus, methods, findings, and other completed aspects of the research process. If that is their only function, those interested in the cognitive activities of scientific inquiry may regard diagrams as epiphenomenal. We contend that in fact scientists use diagrams as tools throughout the research process. In this paper we focus in particular on the distinctive functions served by diagrams in three aspects of mechanistic research:

(1) delineating the phenomenon to be explained;
(2) identifying explanatory relations (relations between variables that are relevant to explaining the phenomenon);
(3) constructing and revising a mechanistic explanation of the phenomenon.

The term *diagram* does not have clear boundaries. Its etymology suggests a very inclusive meaning—any visuospatial representation—which would cover virtually all

A. Abrahamsen · W. Bechtel (✉)
University of California, San Diego, USA
e-mail: bechtel@ucsd.edu

of the figures in a scientific paper including photographs, flow charts of a procedure, and line drawings of an experimental apparatus. Here, though, we focus more narrowly on those diagrams serving the epistemic functions most relevant to mechanistic explanation, which includes graphs and relatively abstract figures but usually excludes drawings and photographs. We distinguish these three types, which correspond to the three aspects of mechanistic research just noted:

(a) Phenomenon diagrams help delineate the phenomenon of interest, often taking the form of a graph depicting the relation between two or more variables;
(b) Explanatory relations diagrams almost always take the form of a graph depicting the relation between two or more variables, at least one of which is not in (a) but may contribute to its explanation through its linkage to a part or operation in (c).
(c) Mechanism diagrams provide a visuospatial representation of the organized parts and operations of a mechanism, which may explain the phenomenon in (a).

This project has benefited from several existing strands of cognitive science research on diagrams. Elucidating how diagrams convey information differently than text was pioneered by Larkin and Simon (1987) and addressed most comprehensively by Tversky (2011), who has emphasized the use of space, icons, and what she calls *glyphs* (simple shapes and lines). Hegarty (Hegarty and Just 1993; Hegarty 2011) has investigated the use of diagrams as tools for reasoning; for example, displaying a diagram of a pulley, she tracks eye movements as people judge the truth of sentences about its expected movements. Cheng (2002, 2011) has developed innovations in the design of diagrams and shown that they improve students' mastery of technical subjects such as electric circuitry and probability theory. In case studies of how specific diagrams were used as tools for the cognitive activities of scientists—physical scientists in particular—Cheng and Simon (1995) focused on Galileo and Nersessian (2008) on Maxwell. We too examine diagrams as tools for scientists but focus on those pursuing mechanistic explanations, which predominate in biology. The most relevant previous work is that of Gooding (2004, 2010) on the role of diagrams in reconstructing extinct organisms. He showed how scientists developed representational formats that enabled their visual processing capacities to see relevant patterns—for example, in diagrams spatially displaying the organized parts of a reconstructed organism.

This project is situated at the nexus of work on diagrams in cognitive science and work on explanation in philosophy of science. The latter is in flux. Mechanistic explanations were pursued by biologists throughout the 19th and 20th centuries. However, they were little discussed in the dominant approaches to philosophy of science in the 20th century, which emphasized derivation from laws as the primary explanatory activity (Hempel 1965; Nagel 1961). Salmon (1984) advanced an influential alternative perspective that focused on causal relations; although often referred to as causal/mechanical, it did not incorporate biologists' focus on the organized parts and operations that compose a mechanism. Now a newer cohort of philosophers, including Bechtel and Richardson (1993/2010), Bechtel and Abrahamsen (2005), Craver (2007), Glennan (1996) and Machamer, Darden, and Craver (2000) are focusing on this kind of mechanistic explanation in biology. Most recently, some have incorporated the

increased role of computational modeling of the dynamics of such mechanisms (Bechtel and Abrahamsen 2010, 2012b) Many are attending as well to the scientists' epistemic commitments and cognitive processes (e.g., Burnston 2013; Burnston et al. 2014; Craver and Darden 2013). Another perspective is offered by philosophers of science who have embraced cognitive science research on distributed cognition (Giere 2002; Osbeck, Nersessian, Malone, and Newstetter 2010). On these accounts, cognitive tasks are distributed across not only agents but also artifacts—which would prominently include diagrams for the tasks of science.

We will use the research field of circadian rhythms as an exemplar throughout this paper, since it provides an especially fruitful specific case in which to examine how diagrams serve as tools for scientists more generally. As the name suggests (*circa* = about + *dies* = day), circadian rhythms are oscillations with a period of approximately 24 h. They are generated endogenously within organisms, but importantly, are entrainable to the day-night cycle in the 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 can be observed in a broad range of physiological and behavioral activities. A major advance was the discovery that the underlying mechanism was a 24-hour molecular clock: intracellular oscillations in the expression of certain genes are responsible for the observed oscillations in metabolism, body temperature, alertness, and numerous other measures. To facilitate discussion without having to introduce too much biological detail, we will focus on the two most-studied variants of the molecular clock, those in fruit flies and mice. Because these mechanisms contain many homologous components that share the same name, we will not emphasize the species differences.

## 2 Phenomenon Diagrams

We begin with diagrams that present the target of explanation—the phenomenon to be explained—in a visual form that supports the scientist's ability to detect salient patterns. In pointing to 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 designate as phenomena to be explained. Many of these cognitive activities take advantage of graphs or other external data displays that help our internal visual information processing system to pick up patterns. The scientist may posit a phenomenon when the same general pattern is obtained repeatedly under appropriate conditions. When a mechanism is proposed to explain that 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.

Several diagram formats have been advantageous for identifying and thinking about circadian phenomena; we will discuss three. We begin with line graphs, the most

ubiquitous format, in which values of a variable of interest can be plotted against time. We then turn to two other formats, actograms and phase response curves, which were developed by circadian researchers to render particular phenomena visually accessible.

*Line Graphs* Body temperature is one among many physiological variables that exhibit a circadian pattern. By recording a person's temperature at regular intervals and plotting these data points on a line graph, we can use our visual pattern recognition capabilities to immediately recognize the oscillatory pattern in which temperature gradually rises each day and drops each night. Discerning the same pattern from a table of numbers requires slower, less compelling cognitive processes.[1] The advantages of the line graph format are illustrated in Fig. 1, in which each data point (small square) indicates one person's body temperature 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 daily oscillations in body temperature. Superimposing light and dark bars on the abscissa's timescale and shading the corresponding areas of the graph are ways to visually highlight the pattern of interest: body temperature dropping towards and during the night and increasing 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.

Although a line graph such as this shows that a variable of interest exhibits circadian rhythmicity, it does not visually convey certain aspects of the rhythms that circadian researchers regard as critical, such as whether there are changes in the exact duration of each cycle (its period) or in the timing of the onset of each cycle (its phase). They therefore have developed certain specialized formats, including actograms and phase response curves, that make such nuances visually apparent.

*Actograms* An actogram is particularly useful for showing a pattern of activity over the course of a day and how that pattern changes (or not) on each subsequent day. Its basic elements are short vertical marks, each representing one occurrence of a designated act (in studies using mice, for example, one turn of a running wheel). The marks are entered left to right on a timeline representing at least one 24-hour period ("day"—typically 1 day-night cycle). The timelines for successive days are stacked from top to bottom, making it easy to trace daily activity patterns across many days and 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 locomotor behavior for one mouse. For the first few days the researchers provided a normal daily Light-dark cycle (LD), as specified by the light and dark portions of the bar at the

[1] The importance of the visual presentation in a graph can be appreciated by considering the study that first established 37 °C (98.6 °F) as normal mean body temperature. Having collected multiple recordings per day from over 25,000 individuals, Wunderlich (1868) noted oscillations of over 1 ° C between a low in the early morning hours and a peak in the afternoon. But he reported these results using summary tables, in which oscillations could be discerned only via effortful, nonvisual processing; this likely was one reason that most subsequent researchers cited him only for establishing the "normal" human temperature, not its daily oscillations.

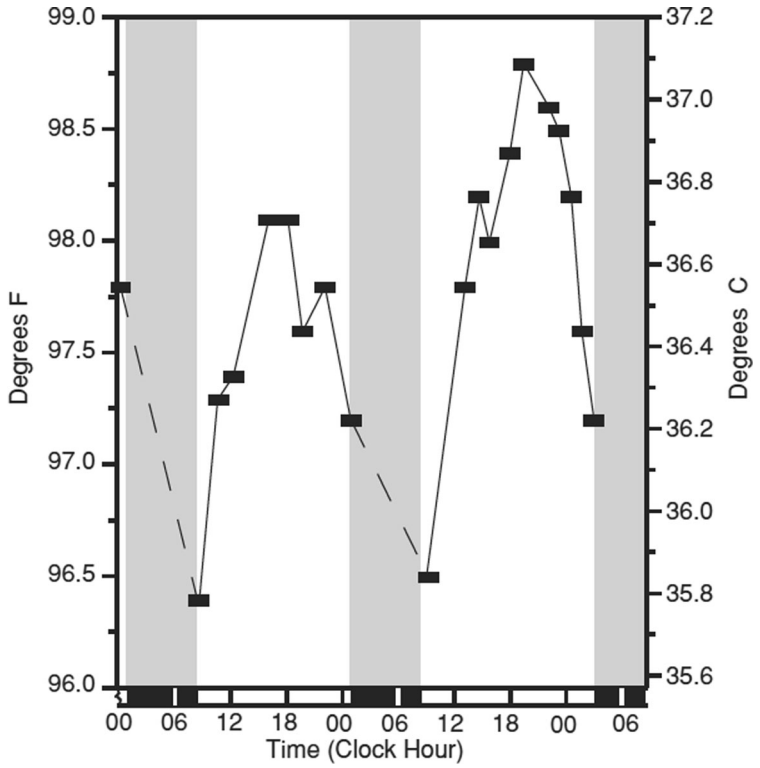


**Fig. 1** Line graph from Koukkari and Southern (2006) showing the circadian oscillation in body temperature for one person across 48 h

top. On the remaining days the mouse was in constant darkness (DD), except for a brief light pulse on 1 day, as indicated by the arrow labeled LP. The timeline at the bottom extends 48 h, not 24 h, because like many other actograms this one is *double plotted*: each day's data is plotted not only below the previous day's data but also (redundantly) to its right. This convention was developed to make it easier to detect patterns, especially by not cutting off an activity phase that straddles the 24th hour. It is visually obvious that different patterns are obtained in the LD vs. DD conditions. In the normal LD condition, this nocturnal animal's activity begins at the onset of darkness and continues until shortly before "dawn." But once in the DD condition (constant darkness, also known as *free-running*), its activity begins somewhat earlier each day, increasingly intruding into what previously had been the hours of light. From these contrasting patterns in the actogram it can be inferred that the period of the mouse's endogenous cycle (revealed by DD) is a bit less than 24 h, and it is entrainment to light in the external environment that stretches the normal (LD) period to 24 h.

Entrainment to light (or to certain other signals, such as a change in temperature) is a central concern in circadian research. One tactic is to interject a single, isolated pulse of light into the DD condition. The actogram in Fig. 2, in addition to visually conveying the activity patterns for LD and DD, shows the effect of one such light pulse on the onset of activity: substantially delayed on the first relevant day, but partially recovered the next day and, resuming the usual DD pattern, a bit earlier each subsequent day.

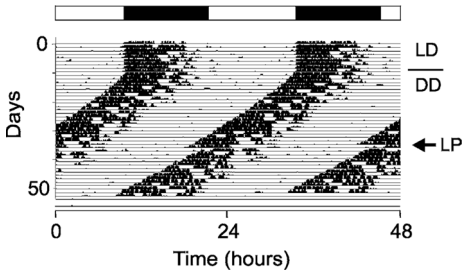


**Fig. 2** Example from Lowrey and Takahashi (2004) of a contemporary actogram. It makes apparent how the activity of a mouse is entrained to light when it is under a light-dark cycle during the initial days (LD), free runs with a period somewhat less than 24 h when in constant darkness (DD), and is only briefly affected by a light pulse (LP)

*Phase Response Curves* In everyday life the capacity for entrainment is what enables us to adjust, albeit slowly, when traveling across time zones or experiencing changes in the amount of daylight at different times of year. While an actogram provides a way to show that a single light pulse can advance or delay the next activity phase, circadian researchers wanted to know more specifically the effect of the timing of the light pulse on the extent of advance or delay. They ran the necessary experiments and developed a specialized type of line graph, the phase response curve, to visually display the quantitative findings.

To obtain Fig. 3, for example, hamsters were first maintained in an LD condition for 7 days and then switched to DD (constant darkness) for another 7 days. On day 15, the researchers provided a 60-minute pulse of light to each hamster at its assigned time within the 24-hour timeline on the abscissa and then recorded when its next activity phase began. The shift in that onset time (relative to the mean activity onset time for the 7 baseline DD days) is shown on the ordinate. Zero shift (horizontal line) indicates no effect of the light pulse; positive values indicate a phase advance and negative values a phase delay. (Note that the timeline is on *circadian time*, in which by convention the time of activity onset for a nocturnal animal is designated as hour 12, the beginning of *subjective night*, and hour 0 is the beginning of *subjective day.* The duration of 24 h of circadian time in this example is a bit less than 24 h of clock time, and each hour is 1/24 of that duration.)

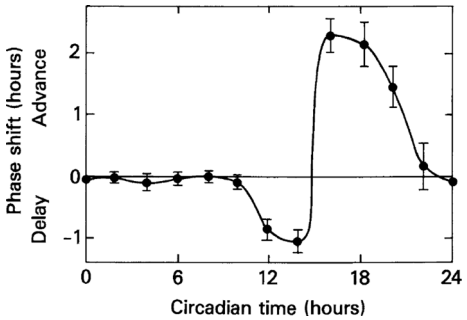


**Fig. 3** Example (Takahashi, DeCoursey, Bauman, and Menaker 1984) of a contemporary phase response curve for a nocturnal species. The time at which a light pulse is delivered is shown on the abscissa and the extent of phase delay or advance is shown on the ordinate

Each data point in the phase response curve is the shift resulting from a light pulse at the indicated circadian time, averaged across the six hamsters tested at that hour. The curve makes clear that a light pulse delivered during subjective day (hours 0–10) has **little effect** on when the next activity phase begins—which is not surprising, since this is when light would have been expected but for the switch to constant darkness. In contrast, a light pulse during the first 2 h of subjective night **delays** activity onset. Again this is not surprising, since the light would signal either that the hamster's circadian clock was running too fast or that, as in spring, the period of daylight was lengthening. The phase delay is evidence of appropriate entrainment. Finally, a light pulse imposed later in the night has the opposite effect: it **advances** the onset of the next activity phase. Thus, the hamsters are interpreting these pulses as signaling an earlier than expected dawn (rather than an extension of dusk), and are entraining accordingly. The effect is most dramatic at hour 16 and weakens as the pulses come closer to the anticipated start of subjective day—a finding that is easy to spot in the phase response curve due to its quantitative precision and good design.

The three diagrams discussed in this section were designed to provide a visual display of the overall phenomenon of circadian rhythmicity (Fig. 1) and of more specific circadian phenomena, especially the shorter period of endogenous cycles (Fig. 2) and the effect of the phase of the endogenous cycle on its entrainment to external light (Fig. 3). Each of the diagram formats used—line graph with light bar, actogram, and phase response curve—is the product of considerable revisions and tweaks to earlier versions (as discussed in detail in Bechtel and Abrahamsen 2012a). Thanks to the good design of today's formats, anyone familiar with them has ready access via visual pattern recognition to key circadian phenomena.

## 3 Explanatory Relations Diagrams

Explanatory relations diagrams are visually indistinguishable from many phenomenon diagrams, in that they display the relations between two or more variables using line graphs or other graphical formats. What makes a particular diagram explanatory is that one or more of its variables is not among those portraying the phenomenon but is causally linked to it—often due to its role in an existing or emerging mechanistic explanation. The notion of an explanatory relation has not previously been recognized in philosophical discussions of explanation. However, the theories, models, or mechanistic accounts that are the focus of such discussions came about when scientists did the empirical research required to find relevant explanatory relations and the mental work of pursuing their implications. Explanatory relations are at the center of actual scientific practice.[2]

As in the case of phenomena, scientists commonly plot raw data or summary measures of data in graphs so that visual perception can be exploited to grasp the pattern that gives the explanatory relation its specific form. This often involves

[2] Our frequent collaborator, Daniel Burnston, first called our attention to explanatory relations and has led our research group's initial consideration of how explanatory relations diagrams figure in scientific practice. The specific construals and applications in this section are ours.

diagramming the same or related data in multiple ways, each designed to enable seeing the relation differently or seeing a different relation. In this section we present two explanatory relations diagrams that provide exemplars of this practice. Notably, both hone in on molecular genetics as the most promising level from which to find explanatory relations to circadian phenomena at the behavioral level. The implications (as pursued in the next section) would include identifying parts of the molecular mechanism responsible for circadian rhythms in locomotor and other behaviors. The second example brings in as well the cellular level, since it focuses on interactions between the molecular mechanisms in different neurons.

Our first example is drawn from the first study that identified (and named) an important part of the molecular mechanism responsible for circadian rhythms: the gene *period*. Starting with a large candidate pool of mutant strains of fruit flies, Konopka and Benzer (1971) identified three strains that produced abnormal circadian rhythms, based on the timing of their eclosion (emergence from the pupa) compared to normal fruit flies. Since this is a once-in-a-lifetime event, rhythmicity was determined by counting the number of flies in a population that ecloded per hour as a function of time across 4 days. Fruit flies normally eclode in early morning, and Konopka and Benzer verified that this was the peak time in their normal population even under free-running conditions (constant darkness, temperature, etc.)—although with some spread across the day, presumably due to the lack of light cues for entrainment. This is shown in the top left line graph (A) in Fig. 4. Comparing the distributions for the three mutant strains, one each shows loss of rhythm (B), a shortened period of ~19 h (C), and a lengthened period of ~28 h (D). Note that this finding involves the two quantitative variables that were graphed plus a third variable, mutation status, that takes four discrete values and was manipulated by the researchers. The explanatory relation is the relation of that variable to the pattern exhibited on the first two variables, most notably, how each of the mutant graphs B-D differs from that of the normal flies in graph A.

Konopka and Benzer traced all three mutations to the same gene locus, and named the gene at that locus *period*. Further, they made the mechanistic inference that *period* is part of a molecular clock responsible for the circadian rhythmicity of eclosion in fruit flies. This leaves open the question of whether *period* is part of a central clock or merely of a specialized clock for the regulation of eclosion. Is the rhythmicity of locomotor activity, for example, explained by a different specialized clock built from other genes? To begin to answer this, Konopka and Benzer measured locomotor activity under free-running conditions in flies from the same four populations. The patterns of results differed across the groups in the same way for locomotor behavior as for eclosion, providing support for a central clock.

Konopka and Benzer used the actograms shown on the right in Fig. 4 to make these locomotor patterns visually compelling. In panels A and B they used a standard 24-hour day so that, with double-plotting, each line depicted activity across 48 h. Since in the free running condition the normal fly exhibited an endogenous period of 24.5 h, the activity shown on each subsequent line is slightly offset, producing a slight left-to-right diagonal. The arrhythmic fly in panel B shown no regular pattern of activity. Had the researchers used same 48-hour timescale for the short-period mutant, each day's activity would have begun much earlier, resulting in an extreme right-to-left diagonal. Instead, in panel C they employed a day of 19 h, yielding a 38-hour timescale (two 19-

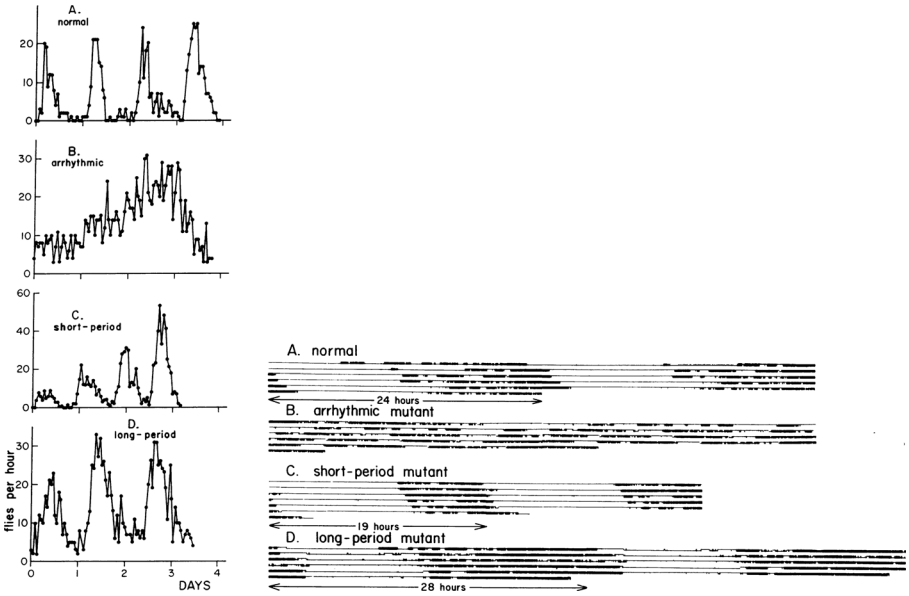


**Fig. 4** *Left*: *Line graphs* from Konopka and Benzer (1971) showing circadian rhythms of eclosion from a population of normal flies (**a**) and three mutant populations (**b**–**d**). *Right*: actogram showing the activity periods of normal flies (**a**) and the three mutant populations (**b**–**d**)

hour days). Since the actual period was 19.5 h, the actogram exhibits a slight left-to-right diagonal similar to that in panel A. Likewise, for the long-period mutant they employed a 56-hour timescale (two 28-hour days). Since the actual period was 28.6, the diagonal is similar to that for the normal fly (vs. an extreme left-to-right diagonal on a 48-hour timescale). By adapting the timescales in this way the active phases stacked nicely across days, so the short periods looked short at a glance, the long periods looked long at a glance, and the half-hour discrepancies for all except the arrhythmic all produced the same slight diagonal.

Our second example brings out even more clearly how researchers often utilize a variety of diagram formats to understand explanatory relations. This research focused on relations between different neural cells in a mammalian brain region that is specialized for coordinating circadian rhythms across the organism, the suprachiasmatic nucleus (SCN). Individual neurons in the SCN have been shown to maintain circadian rhythms but with varying periods. A regular circadian rhythm is generated only by synchronizing the activity of these individually oscillating neurons, and (Ciarleglio, Gamble, Axley, Strauss, Cohen, Colwell, and McMahon 2009) investigated the role of a particular molecule released by some SCN cells, vasoactive intestinal polypeptide (VIP), in achieving the synchrony that is important in producing circadian behavior. Specifically, in Fig. 5 the researchers compared six mice that differed in their mutation status: $VIP^{+/+}$ had two normal copies of the VIP gene; $VIP^{+/-}$ had one normal and one deleted copy, and $VIP^{-/-}$ had both copies deleted. Within each mutation status, one mouse had been maintained under a light-dark cycle (LD) and the other in total darkness (DD).

In an earlier figure (not reproduced here), the researchers first provided actograms of running wheel activity to show how these genetic and environmental conditions affected behavioral cycles. More indirect measures were needed to determine how they affected the molecular clocks within SCN neurons. For tractability these researchers focused on just one clock component: a mammalian version of the *period* gene first identified in fruit flies by Konopka and Benzer. As the expression activity of the two *period* genes in the neuron's nucleus oscillate, the result (at some delay) is similarly oscillating concentrations of PER proteins in the neuron's cytoplasm. Ciarleglio et al. attached a green fluorescent reporter gene to *period* such that the expression activity of the two genes was yoked. In consequence, oscillations in PER protein concentrations could be indirectly measured by visually tracking the fluorescent proteins. Specifically, values of relative fluorescence intensity over time were obtained with a special camera directed at SCN tissue slices (which, though removed from the rest of the SCN, continued to function).

The line graphs in Fig. 5 display these values for the slices from each of the six mice. The overall fluorescence fluctuations across 96 h are in Panel A, broken down to show the variations across individual neurons in Panel B. It can be seen that the molecular clock within each neuron continues to produce oscillations under all conditions, but that the clocks become desynchronized within 2 days in the absence of VIP (strain $VIP^{-/-}$) whereas in other conditions the oscillations remained fully synchronized ($VIP^{+/+}$) or partially synchronized ($VIP^{+/-}$).

Panel C represents the Day 1 data using a different format, the Rayleigh plot, in which the 24 h of a single day are arranged in a circle rather than horizontally. The time

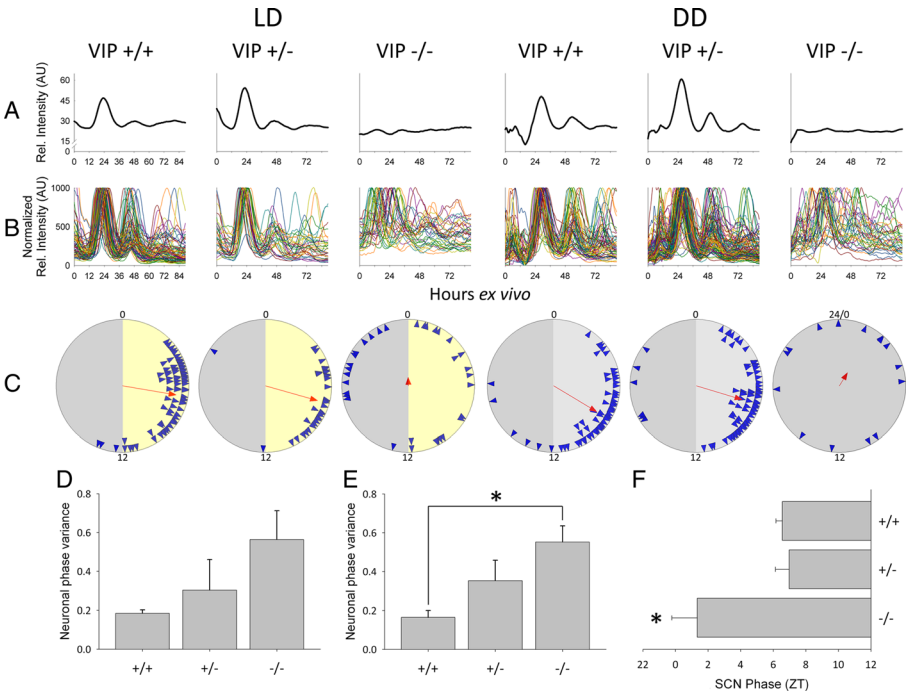


**Fig. 5** Multiple diagram formats Ciarleglio et al. (2009) employed to identify a family of explanatory relations between VIP and the synchronization of SCN neurons

at which each detectable neuron in the slice reached 50 % of its maximum fluorescence is marked by a blue arrowhead.[3] The distribution of these arrowheads presents a different way of visualizing the contrast between synchronized and desynchronized clock activity in neurons. The red arrow in each plot shows the direction and strength of the overall vector characterizing that slice (which is very weak for the two $VIP^{-/-}$ mutants). Finally, panel D uses yet another format, bar graphs, for a crisper depiction of the Panel B finding that the loss of VIP results in more variable phase timing. The last bar graph informs us that this also results in a 6-hour phase advance on average. Although the diagrams in Panels A-D all draw on the same data, they identify different relations in those data that the researchers viewed as explanatorily relevant at the neural and molecular levels. More specifically, VIP is not a component of the clock in each neuron but is causally involved in the synchronization of clocks across neurons.

Each diagram in this section displays evidence that a manipulated variable was causally related to variables characterizing the phenomenon of interest: an explanatory relation. Such relations can take a variety of forms (differing in number of variables involved, whether they are discrete or quantitative, etc.), as can the diagrams conveying them (actogram, line graph, bar graph, Rayleigh plot, etc.). The diverse options can be advantageous for obtaining different perspectives on the same data. Most important, though, is what they have in common: each diagram makes the explanatory relation accessible via visual pattern recognition.

Researchers often stop here, satisfied to have identified a variable that in some way is causally related to the phenomenon of interest and hence explanatorily relevant to it. Our interest is in those who take the next step, inferring from the specific causal relation either the first sketch of the mechanism responsible for the phenomenon or some addition or refinement to an existing mechanistic explanation. The examples in this section point to the *period* gene as a part in the molecular clock mechanism, and VIP as a messenger between the clocks in different neurons. In the next section we examine another researcher's diagram showing *period* and possible associated parts and operations as well as a later, more advanced version of the molecular clock mechanism.

## 4 Mechanism Diagrams

We turn finally to mechanism diagrams: those using icons or labels to represent various parts of a mechanism and arrows or other devices to represent operations through which parts are transformed or affect other parts. Although limited to the two dimensions of the page, researchers sometimes use spatial relations to directly represent actual spatial locations of parts and operations. More often, they use the space to cluster parts and operations that affect each other, so as to form a conceptual perspective on the organization of the mechanism and how it works. Such diagrams support scientists' reasoning as they plan experiments to help fill in missing parts, explore alternative configurations, mentally simulate the flow of activity through the mechanism, at least to rule out versions that cannot work, or in many other ways seek to verify or improve

[3] At those times of day with the most molecular clock activity, arrowheads are stacked further into the circle as necessary to keep them distinct. The number of arrowheads per plot is not of interest; it simply indicates how many neurons in the slice fluoresced.

the account. As more experiments are completed, revealing additional explanatory relations, more elaborate diagrams bring together their implications in user-friendly formats.

We present two examples of mechanism diagrams in Fig. 6. Both were developed in the process of achieving a mechanistic explanation of circadian rhythms in fruit flies. Hardin, Hall, and Rosbash (1990), building on Konopka and Benzer's discovery of the gene *period* (*per*), were the first to demonstrate 24-hour oscillations in the concentrations of *per* mRNA and (at some delay) PER protein. To account for this they proposed that PER figured in some sort of feedback process whereby the greater its concentration, the more it inhibited its own further production. It was known that a negative feedback loop was the type of organization that could generate oscillations, and that with sufficient delays and non-linearities these oscillations could be sustained indefinitely.

Hardin et al. therefore offered the diagram on the left side of Fig. 6, not as a firm proposal but rather as one that laid out in a common space the possible ways such a negative feedback loop might be constructed. There were three possible origins of the feedback: the PER protein itself (X); an unidentified biochemical product of PER (Y); or some behavior (Z) of the organism that in some way relied on PER. Moreover, there were two possible targets of the negative feedback, either of which would inhibit protein production: *per* and its transcription into mRNA; or mRNA and its translation into PER. The question marks on these alternative paths are a notable feature, used in many mechanism diagrams in biology. They 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, pending empirical research. As results pruned some of the paths, a diagram like this could serve as a dynamic tool in achieving an account of the molecular mechanism responsible for a phenomenon of interest (here, circadian rhythmicity).

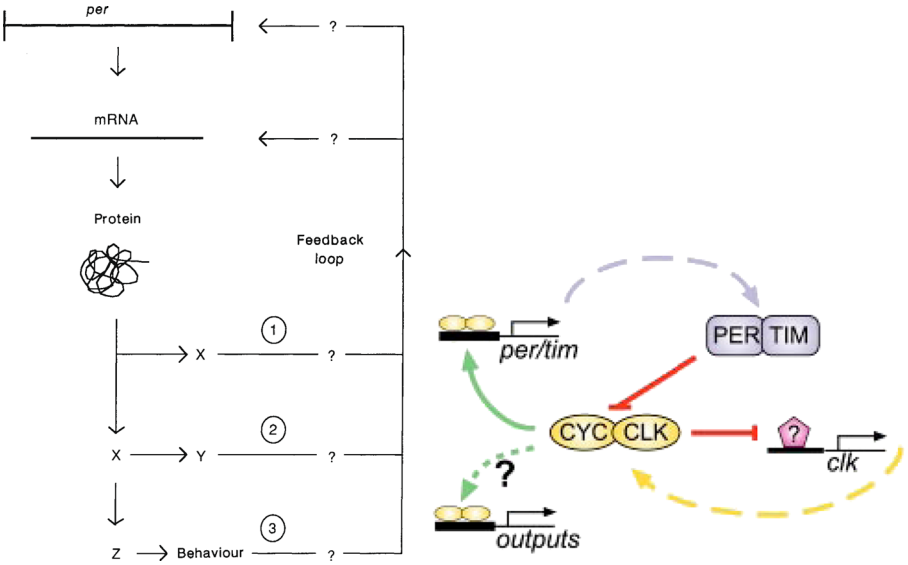


**Fig. 6** *Left*: Hardin, Hall, and Rosbash's (1990) mechanism diagram proposing a transcription-translation feedback loop for generating circadian rhythms in fruit flies. *Right*: Harmer, Panda, and Kay's (2001) diagram showing how the understanding of the fruit fly clock mechanism had developed over the following decade. Note the prominent use of question marks in both figures

The diagram on the right side of Fig. 6 was produced a decade later by Harmer, Panda, and Kay (2001). By that time researchers had established a substantial network of explanatory relations from which answers and elaborations of the mechanistic account could be inferred. Those incorporated in this diagram are as follows.[4] (1) A number of additional clock genes and their proteins had been identified, including the proteins TIMELESS (TIM), CYCLE (CYC), and CLOCK (CLK). (2) The origin of the negative feedback did not exactly correspond to any of Hardin et al.'s three alternatives: it was a dimer (two-part compound), PER:TIM. (3) The target was also more complex than expected. It turned out that *per* and *tim* could be transcribed only when another dimer, CYC:CLK, was bound to their promoters. During the negative-feedback phase of the cycle, PER:TIM blocked this binding (rather than acting directly on *per*, *tim*, or their mRNA). (4) There was at least one additional negative feedback loop. Harmer, Panda and Kay included these now-understood parts and operations in their diagram without question marks. Instead, they used question marks to denote two remaining gaps in knowledge: what bound to the promoter of *clk*, and what served as the output that enabled this molecular clock mechanism to regulate expression of other genes in the organism.

The two mechanistic diagrams in Fig. 6 provide just a glimpse of the diverse formats, styles, and uses of mechanistic diagrams not only by circadian researchers but across a broad range of fields. They are ubiquitous in laboratories: stacked on desks, inserted into lab notebooks, drawn on whiteboards with warnings not to be erased, and so forth. These diagrams support scientists' cognitive activities as they develop mechanistic explanations and revise them in the light of new findings. Some are used as well for communication and interaction within the research group or in presentations. A very small number get published. (See Burnston et al. 2014, for a case study in which more than a dozen working versions preceded the published version of an innovative circadian diagram.)

One impetus to diverse formats is the challenge of representing time in the static medium of a mechanism diagram. None of the existing solutions (see Bechtel et al. 2014) can be illustrated in this paper, but one example is a multipart figure with the same basic diagram repeated but modified to show its different states at four different times of day. Alternatively, if a computational model exists that captures the dynamics of a particular mechanistic account, the model can be anchored to a diagram of the mechanism by appending appropriate variables to its parts and operations. Typically the model will be a system of differential equations, which indirectly brings time to the diagram and may provide a test of the mechanism's ability to produce the phenomenon. The increasing attention to this kind of coordination between a basic mechanistic account and a computational model of the mechanism's dynamics is such a major advance that we regard it as a distinct type of enterprise, *dynamic mechanistic explanation*. For discussion of this approach and the distinctive types of diagrams developed for it, see Bechtel and Abrahamsen (2010, 2012b).

[4] Many diagrams of this type, with the addition of a curved line representing the nuclear membrane, incorporate the fact that *per* mRNA is transported across the membrane into the cytoplasm, where it is translated into PER, which then dimerizes with TIM. To initiate negative feedback, the dimer must be transported back into the nucleus. Together with other time-consuming operations, these achieve a 24-hour cycle.

**Acknowledgments** We gratefully acknowledge the support of National Science Foundation grant 1127640 and the numerous contributions of our collaborators Daniel Burnston and Benjamin Sheredos.

## References

Bechtel, W., and A. Abrahamsen. 2012a. 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., and A. Abrahamsen. 2012b. Thinking dynamically about biological mechanisms: Networks of coupled oscillators. *Foundations of Science* 1–17.

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

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

Bechtel, W., D. Burnston, B. Sheredos, and A. Abrahamsen. 2014. Representing time in scientific diagrams. *Proceeding of the 36th Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Bogen, J., and J. Woodward. 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.

Burnston, D. C., B. Sheredos, A. Abrahamsen, and W. Bechtel. 2014. Scientists' use of diagrams in developing mechanistic explanations: A case study from chronobiology. *Pragmatics and Cognition*.

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., and H.A. Simon. 1995. Scientific discovery and creative reasoning with diagrams. In *The creative cognition approach*, ed. S.M. Smith, T.B. Ward, and R.A. Finke, 205–228. Cambridge: MIT Press.

Ciarleglio, C.M., K.L. Gamble, J.C. Axley, B.R. Strauss, J.Y. Cohen, C.S. Colwell, and D.G. McMahon. 2009. Population encoding by circadian clock neurons organizes circadian behavior. *Journal of Neuroscience* 29: 1670–1676.

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

Craver, C.F., and L. Darden. 2013. *In search of mechanisms: Discoveries across the life sciences*. Chicago: University of Chicago Press.

Giere, R.G. 2002. Scientific cognition as distributed cognition. In *The cognitive bases of science*, ed. P. Carruthers, S. Stich, and M. Siegal. Cambridge: Cambridge University Press.

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

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., J.C. Hall, and M. Rosbash. 1990. Feedback of the *Drosophila period* gene product on circadian cycling of its messenger RNA levels. *Nature* 343: 536–540.

Harmer, S.L., S. Panda, and S.A. Kay. 2001. Molecular bases of circadian rhythms. *Annual Review of Cell and Developmental Biology* 17: 215–253.

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

Hegarty, M., and M.A. Just. 1993. Constructing mental models of machines from text and diagrams. *Journal of Memory and Language* 32: 717–742.

Hempel, C.G. 1965. Aspects of scientific explanation. In *Aspects of scientific explanation and other essays in the philosophy of science*, ed. C.G. Hempel, 331–496. New York: Macmillan.

Konopka, R.J., and S. Benzer. 1971. Clock mutants of *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences of the United States of America* 89: 2112–2116.
Koukkari, W., and Southern, R. N. 2006. *Introducing biological rhythms*. New York: Springer.
Larkin, J.H., and H.A. Simon. 1987. Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science* 11: 65–99.
Lowrey, P.L., and J.S. Takahashi. 2004. Mammalian circadian biology: Elucidating genome-wide levels of temporal organization. *Annual Review of Genomics and Human Genetics* 5: 407–441.
Machamer, P., L. Darden, and C.F. Craver. 2000. Thinking about mechanisms. *Philosophy of Science* 67: 1–25.
Nagel, E. 1961. *The structure of science*. New York: Harcourt, Brace.
Nersessian, N. 2008. *Creating scientific concepts*. Cambridge: MIT Press.
Osbeck, L.M., N. Nersessian, K.R. Malone, and W.C. Newstetter. 2010. *Science as psychology: Sense-making and identity in science practice*. Cambridge: Cambridge University Press.
Salmon, W.C. 1984. *Scientific explanation and the causal structure of the world*. Princeton: Princeton University Press.
Takahashi, J.S., P.J. DeCoursey, L. Bauman, and M. Menaker. 1984. Spectral sensitivity of a novel photoreceptive system mediating entrainment of mammalian circadian rhythms. *Nature* 308: 186–188.
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.