

# Diagramming Phenomena for Mechanistic Explanation

**William Bechtel (bechtel@ucsd.edu)**

Department of Philosophy, University of California, San Diego  
La Jolla, CA 92014 USA

**Adele Abrahamsen (aabrahamsen@ucsd.edu)**

Center for Research in Language, University of California, San Diego  
La Jolla, CA 92014 USA

## Abstract

As part of an inquiry into how diagrams figure in scientific practice, we examine diagrams that represent phenomena involving circadian rhythms. Different diagrammatic formats are developed and revised over time to best represent different phenomena for which explanations will be sought. Some diagrams are less transparent than others, so learning is often required in order to see the information conveyed.

**Keywords:** Diagrams; Graphs, Mechanistic explanation, Visual representation; Circadian rhythms.

## Introduction

The notion of representation covers a lot of territory in cognitive science, encompassing both internal and external encodings of information and a variety of formats. Cognitive scientists have long focused on language-like internal representations, with some dispute over possible ways they might be supplemented by analog formats. Recent years have brought increased attention to external representations and especially to those incorporating analog formats—that is, diagrams. For example, Hegarty (2004) has shown how individuals perform simulations with diagrams in solving problems, and Cheng (2011) has explored how alternative diagramming techniques can foster learning. However, except for the pioneering analysis by Nesson (2008) of the role of diagrams in Maxwell's discoveries, there has been little investigation of the use of diagrams in science. Almost all scientific papers include diagrams, and readers often focus on these as they navigate a paper. They are well suited not only for displaying instruments, techniques, multistep procedures, and results but also scientific reasoning. Most generally this involves the construction, evaluation and revision of hypotheses but our particular interest is in sciences pursuing mechanistic explanations, notably the life sciences.

The project of explaining a phenomenon by identifying and understanding the mechanism responsible for it has roots in the scientific revolution beginning in the 16th century. Descartes posited that phenomena such as magnetic attraction are generated by the coordinated activities of constituent parts (in his case, hypothesized corpuscles). He applied the idea of contact action between particles to explain not just physical phenomena, but nearly all phenomena exhibited by living organisms. The only exceptions were reasoning and language use, which he attributed to an immaterial mind because he could not conceive of a mechanism capable of constructing novel, semantically appropriate

thoughts or sentences. The idea of mechanistic explanation quickly took root in biology. Although resisted by vitalists, who contended that something beyond physical processes was required for the functions of life, other early investigators of physiological phenomena embraced mechanistic explanations. As their inquiries progressed, researchers expanded the range of operations involved in biological mechanisms beyond Cartesian physical contact. Newtonian forces, chemical bonding, and electrical conductance were among the operations used in explaining such phenomena as metabolism, nerve transmission, and heredity.

Mechanistic explanation was largely overlooked by 20<sup>th</sup> century philosophers of science, who drew from physics the idea that scientists explain phenomena by deriving them from laws (Hempel, 1965). More recently, philosophers focusing on the life sciences have moved the spotlight once again to mechanistic explanation (Bechtel & Richardson, 1993/2010; Bechtel & Abrahamsen, 2005; Machamer, Darden, & Craver, 2000). Typically, life scientists treat the system that generates a phenomenon as a mechanism. They decompose it into parts and operations and then recompose it (conceptually, physically, or mathematically) to arrive at an account of how the coordinated performance of these operations could indeed generate the phenomenon.

Although one may try to describe linguistically the parts and operations of a mechanism and how they interact, often telling a narrative about how each part in succession performs its operation, diagrams generally provide a more useful representational format for conceptualizing and reasoning about a mechanism. Parts may be represented by labels, symbols, or abstract shapes, and the operations by which they interact represented by arrows. Diagrams can illustrate the structural and functional relations between many components and allow viewers to direct their attention successively to different activities that may be occurring concurrently in the mechanism.

The initial step in mechanistic research, though, is delineation of the phenomenon to be explained, and that is where we begin our inquiry into diagrams. Linguistic descriptions of phenomena have been the focus in many philosophical accounts of mechanistic explanation (e.g., “proteins are synthesized by constructing strings of amino acids in the order specified in a sequence of DNA”). However, scientists typically work with much more specific accounts of phenomena, often incorporating numerical values determined in their research. Frequently the numerical data relied upon in char-

acterizing the phenomenon is presented in tables. As Bogen and Woodward (1988) made clear, however, explanations are directed not at the data but rather at the pattern extracted from the data—the phenomenon. Some data patterns can be captured in one or a few equations, such as  $AI / I = k$  (Weber's law). Even when equations suffice, but especially when they do not, scientists turn to diagrams to present the phenomenon. Diagrams turn out to be extremely useful for phenomena that exhibit interesting dynamics—patterns of change over time. Well-known examples include tide tables and EEG recordings.

To gain an understanding of these uses of diagrams in the actual practice of science, we focus on a domain of biology in which dynamics are fundamental: circadian rhythm research. Circadian rhythms are oscillations in activity with an approximately 24-hour cycle. They are endogenously generated but entrained to the timing of the day-night cycle in specific locales at different times of the year. They have been identified wherever sought, not only in animals but also in plants, fungi, and even cyanobacteria. They regulate a vast array of physiological processes (e.g., basic metabolism and body temperature) and behaviors (e.g., locomotion and reaction times in cognitive tasks).

Diagrams, of course, are processed visually. Although visual processing is highly complex (involving nearly half of the cortex in primates; see van Essen & Gallant, 1994), seeing often seems transparent: as we look out into the world, we have the impression that we directly perceive the identity and arrangement of objects in the visual field. The apparent transparency of diagrammatic representations of phenomena is part of their appeal—but this is deceptive. Transparent seeing is often the result of a great deal of learning. There are numerous experiments showing this, using complex scenes or illusory stimuli, but diagrams offer potent demonstrations as well. Some techniques of diagramming are so familiar and straightforward that we readily *see* what the diagram is meant to convey. But other techniques are new to us, and we must go through a process of learning before we see what is presented in the diagram. This is clearly true of scientific diagrams, as we will illustrate in examining what are likely for most readers to be unfamiliar diagrammatic formats developed by circadian researchers.

A related characteristic of diagrams on which we will focus is that scientists develop techniques for diagrammatic representation over time. Sometimes in seeking to represent new phenomena, they can borrow a format that had been developed elsewhere and is already well understood. However, existing formats may not offer the best vehicle for revealing what is significant in the phenomenon or for engaging in further reasoning about it. This drives scientists to develop new representational formats. Like other innovators, scientists typically must apply multiple rounds of revision to their first attempts at novel diagrams, finally achieving a format that meets their cognitive needs. Accordingly, representational formats employed in a science are not static but are developed and changed, and such changes can in

turn alter the cognitive processes of the scientists who construct them.

In this paper we examine several diagrammatic formats that researchers have developed to represent circadian phenomena. We especially focus on how these formats were introduced and revised and show what users of the diagrams must learn in order to interpret them.

## Diagrammatic Representations of Circadian Rhythms

A very familiar way of representing rhythmic oscillation is to employ a Cartesian coordinate system in which time is presented on the abscissa and values of a variable of interest on the ordinate. Many such examples can be found in diagrams of circadian phenomena. Figure 1 shows Aschoff & Wever's (1981) plot of potassium levels in urine samples taken every four hours from six individuals. To make the circadian pattern immediately apparent, they shaded the hours of darkness and connected each individual's data points to yield six superimposed line graphs.

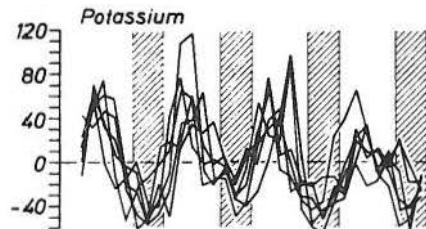


Figure 1. Circadian oscillations in potassium levels for six individuals across four days, measured every four hours (From Aschoff & Wever, 1981).

Such diagrammatic representations make manifest whether the oscillations are regular and sustained (versus damped), the duration of each cycle, and the average amplitude and extent to which it varies across cycles. These can also be compared across diagrams for different individuals (as in Figure 1), variables, or conditions. But there are limitations to the transparency of such diagrams; some types of information are much less perspicuously represented. Accordingly, circadian rhythm researchers have developed other diagrammatic formats, each making particular phenomena manifest.

### Actograms

Even when Cartesian plots incorporate some representation of a reference oscillation (e.g. the light-dark cycles in Figure 1), they are not the ideal display for tracking the variable of interest with respect to those cycles across multiple days. A far more suitable format for this purpose is an actogram, in which stacked horizontal lines represent successive days, and short vertical lines mark each time the variable exceeds some threshold. (Where the vertical lines are dense, some investigators simplify the plot by substituting a solid horizontal bar.) Visual inspection then quickly makes clear any systematic changes across days in the circadian cycles.

The technique of representing activity in actograms appears to have been developed 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 2, left side). While one could compare multiple circular tracings to assess changes or stability over successive days, Johnson introduced the actogram as a better format for this purpose. In essence, he unrolled each circular tracing (one day's data) into a straight line and placed each line below the previous one such that the hours of all days were in alignment. In this first actogram (Figure 2, right side), three sets of lines were presented so as to compare mice from three different environments (greenhouse, from lab to woods, and woods). Within a set, the top line is Day 1, second line is Day 2, and so forth. Comparison was made more precise by running a vertical line at two-hour intervals, beginning and ending at 6 am. A shorter vertical line marked the onset of sunset, a convention not maintained by subsequent researchers. A major virtue of the actogram is that viewers can employ their visual ability to detect differences in the pattern of marks so as to compare circadian rhythms across days or conditions. It can be seen that these nocturnal animals are active primarily at night, with variations between conditions in onset time and in the extent of daytime activity but considerable stability across days within each environmental condition.

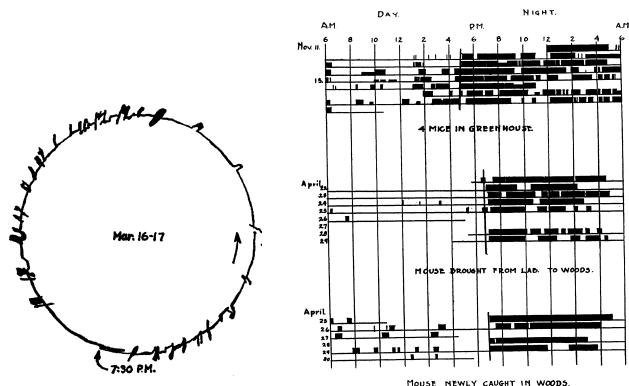


Figure 2. On the left, Johnson's (1926) tracing from a rotating disk illustrating periods of activity by a mouse. On the right, his first actogram in which activity for each day under different conditions is shown on successive lines and the time of sunset is indicated by a vertical line.

In a subsequent investigation (Figure 3) Johnson showed the effects of exposing a deer mouse to different light-dark conditions (as indicated by labels along the right edge rather than by spatially separating conditions): (1) normal light-dark cycle (light during daytime hours, dark at night); (2) constant darkness for several 24-hour periods; (3) reversed light-dark cycle (dark from 8 am to 8 pm, then light from 8 pm to 8 am); (4) again, constant darkness. One can immediately see that the periods of activity showed little change when the mouse was transferred from normal light-dark

cycles to constant darkness, but shifted dramatically when light was reintroduced in reverse: with the dark hours now in daytime, the mouse became active during day rather than night hours. This altered pattern was maintained when constant darkness was reintroduced. Thus, once entrained to a particular light-dark pattern (normal or reversed), mice kept to the same activity cycle when the entrainment stimulus (light) was removed.

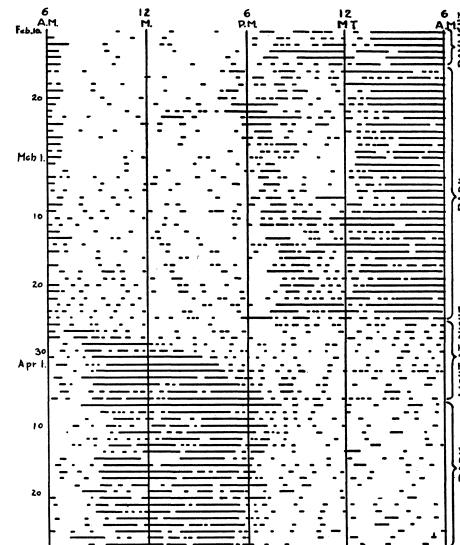


Figure 3. Johnson's (1926) second actogram, which shows a mouse's activity as light-dark conditions were changed as indicated on the right.

Since introduced by Johnson, the actogram has become a standard way of representing circadian activity, especially in animals. While the basic format has been preserved (one line per 24 hours, hours aligned vertically, active times marked along the line), many variations have been introduced to make specific features of circadian behavior more explicit. As just illustrated, chronobiologists are particularly interested in activity under constant darkness (known as *freerunning* behavior). Since this condition eliminates any effects of daily entrainment to sunlight, it can reveal the animal's normal endogenously generated rhythm or (as in Figure 3) the enduring effects of resetting that rhythm via an abnormal light-dark condition. Various conventions have been adopted for conveying lighting conditions visually, rather than by labels along one side as in Figure 3. One is background shading of the actogram across those hours the organism is in darkness (similar to the shading superimposed on the Cartesian plot in Figure 1). Another common convention is horizontal bars at the top or bottom of the actogram, in which white indicates hours of light and black indicates hours of darkness. Thus, in Figure 4 the top bar indicates the normal light-dark condition used as a baseline on days 1-7 and the bar below it indicates that constant darkness was imposed thereafter. It can be seen that once in constant darkness, the mouse's activity begins about a half hour earlier each day. From this it can be concluded that the endogenous period is about 23.5 hours.

A variety of conventions have been developed to indicate temporary changes in conditions. In Figure 4, the gray arrow indicates a day on which a light pulse was presented four hours after activity onset. This not only caused activity to mostly cease for that evening, but also inserted a phase delay the next day into what was otherwise a continuing pattern of phase advance due to constant darkness.

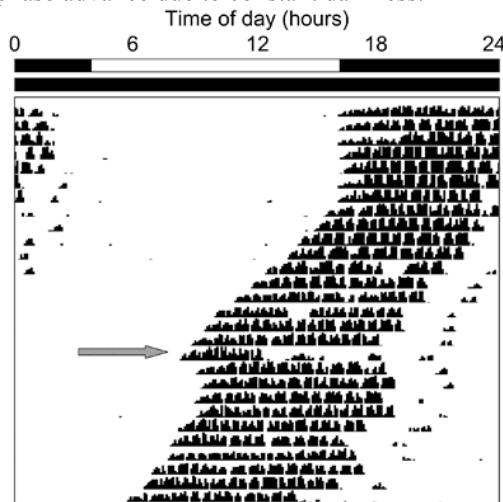


Figure 4. Contemporary actogram in which the top bars indicate a normal light-dark condition for the first seven days and constant darkness for subsequent days. The grey arrow identifies the day a light pulse was administered. (From <http://www.photosensorybiology.org/id16.html>.)

A major innovation in actograms was the introduction of double-plotting—a procedure in which data from the next 24-hour period is plotted not only on the next line but also to the right of the current data. Thus, each line shows data from 48 hours, but the left half of the actogram stacks all 24-hour periods as usual (as does the right half, redundantly). One of the first uses of this technique, by Pittendrigh (1960), well illustrates its particular advantage when activity periods extend across the 24-hour boundary. Pittendrigh was seeking to represent what he called an “after effect”: altered circadian rhythms in the activity of a nocturnal animal after exposure to continuous light. He began with normal light-dark cycles (LD). As shown in the top third of Figure 5, the mouse is inactive when the light is “ON” and becomes active when the light switches to “OFF” (with some subtleties in the data we need not discuss). When he then imposed continuous light (LL), the resultant progressive delay in activity onset indicated the mouse’s day had been stretched longer than 24 hours. After a few weeks, however, it spontaneously shifted back to a nearly 24-hour period. Since double-plotting was new, Pittendrigh marked the divide between the two 24-hour periods with a double vertical line, a convention later dropped as researchers became accustomed to double-plotting. (The convention of indicating lighting conditions with white/black horizontal bars had not yet been adopted, hence Pittendrigh’s ON/OFF markers). The virtue of double-plotting is that one can easily see the full active phase even when it crosses the 24-hour boundary.

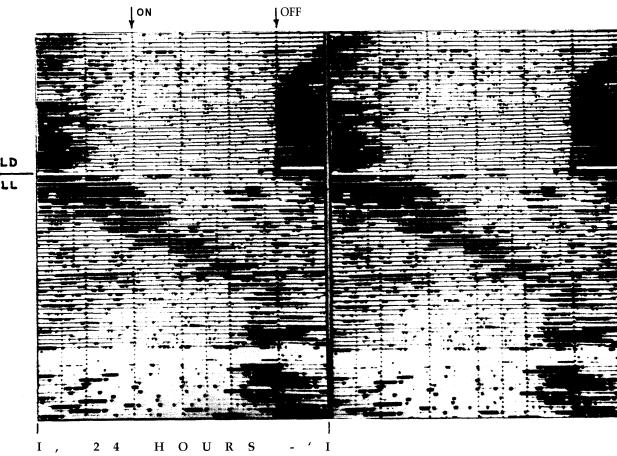


Figure 5. Double-plotted actogram from Pittendrigh (1960), in which data across the 24-hour boundary can be viewed in the middle of its redundant 48-hour display.

Because they provide an effective visual display by which researchers can immediately see variations in patterns of circadian activity, especially in relation to different lighting conditions, actograms have remained an important part of the toolkit for those circadian researchers who use animal behavior measures.

### Phase Response Curves

The circadian cycle can be thought of as a progression through phases, beginning at the time designated hour 0 (typically dawn) and ending at hour 24. The term *phase* may be used with respect to particular points on the curve (the peak and the trough being of particular interest) or for intervals (e.g., subjective day and subjective night under constant darkness). When two different cycles are closely aligned, they are said to be *in phase*. As already noted in the discussion of Figure 4 above, an actogram can show whether and by how much a rhythm is reset following a light pulse (its phase shift). As they explored this phenomenon in the 1950s, circadian researchers soon recognized that the direction and extent of the phase shift depended on the timing of light-dark cycles relative to the animal’s current circadian cycles. Individual instances of resetting could be shown in actograms, but to identify and represent the systematic pattern of resetting, researchers developed what are known as *phase response curves*.

Examining one of the first attempts to represent the effect of light on circadian phase makes evident the challenges in developing an easily interpreted diagrammatic format. Hastings and Sweeney (1958) grew *Gonyaulax polyedra*, a photosynthetic marine dinoflagellate that produces luminescence when disturbed, first under a normal light-dark cycle and then in constant darkness. Next they exposed these organisms to a three-hour pulse of light, varying the time of the pulse so as to determine how much that shifted the time of maximal luminescence. Their results are shown in Figure 6, where hour 0 is the onset of constant darkness. The time of maximal luminescence in control organisms (who were

not exposed to the pulse) is shown by the vertical lines at 7, 31, and 55 hours. The horizontal lines represent organisms exposed to the three-hour pulse at different hours of delay after the onset of darkness (3, 7, 11, . . .). Curves have been fit to data points marked by small triangles, which indicate the time of their subsequent maximum luminescence. The distance of each data point from the nearest vertical line represents the degree of advance or delay. It can be seen by following the horizontal line labeled "23" 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. While this diagram does encode the crucial information, interpreting it takes considerable effort.

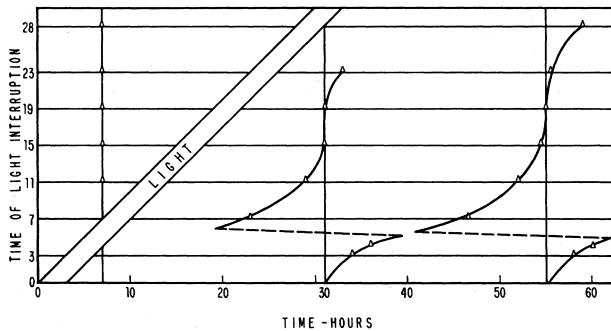


Figure 6. Hastings and Sweney's (1958) diagram showing the changes in peak luminescence of *Gonyaulax polyedra* after 3-hour light pulses. These changes are represented by the distance left or right from the vertical lines.

Shortly thereafter DeCoursey (1960) introduced a different format for representing the same information (Figure 7). 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 an animal's running-wheel activity. The data points indicate the consequent advance or delay 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 (corresponding to the beginning of its subjective night) delays its activity, whereas light 8 to 12 hours later (corresponding to the end of its subjective night) advances its activity. Light during its subjective day (from approximately 12 hours to 0 hours) has no effect. Having represented the phenomenon this way, one can also readily make sense of it—light exposure during subjective day does not indicate a need to reset the phase of one's activity, whereas light at the beginning of subjective night 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, light experienced at the end of the subjective night indicates a need to stop its activity sooner.

DeCoursey's phase response curves quickly became the established means of representing the effect of a stimulus on the phase of an organism's circadian oscillation, although later researchers simplified the abscissa to circadian time (0-

24 hours) and, often, flipped the ordinate so that advances are shown as up and delays as down. Represented in this fashion, as in Figure 8, researchers were able to contrast two patterns of resetting—one producing gradual advances or delays (*Type 1*) versus an alternative (*Type 0*) in which, rather than small advances or delays, at a critical point the organism exhibits a large delay. If this delay is more than 12 hours, it can be seen as a large advance.

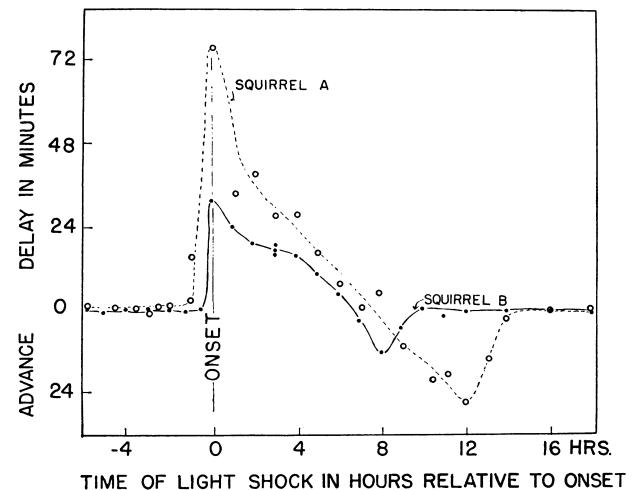


Figure 7. DeCoursey's (1960) phase response curve, which shows the advance or delay of a rat's activity onset for light pulses at different times relative to normal activity onset.

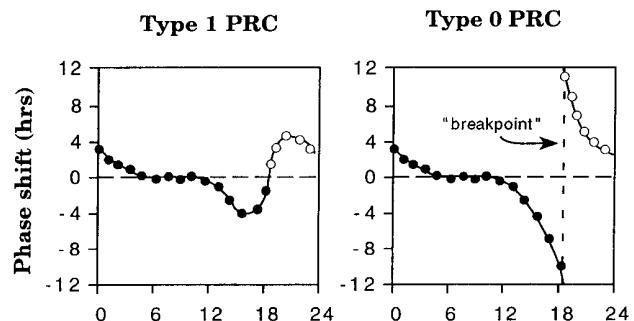


Figure 8. Type 1 and Type 0 phase response curves, with a simplified abscissa (0=dawn). From Johnson (1999).

To facilitate characterization of these two types of resetting, circadian researchers developed yet another diagram format, the phase transition curve, in which the new phase, not the amount of advance, is plotted on the ordinate. Diagonal lines indicate the situation in which the phase does not change. To see what happens in Type 0 resetting, one must plot 48 hours on the ordinate. As seen in figure 9, Type 1 resetting is characterized by a curve that stays very close to the diagonal and so approximates a slope of one (from which the name Type 1 is derived). In Type 0 resetting there is an abrupt jump from one diagonal to another, approximating a slope of 0. A virtue of the phase transition curve is that it makes clear the relation of the new phase to the old, which is not directly displayed in phase response curves.

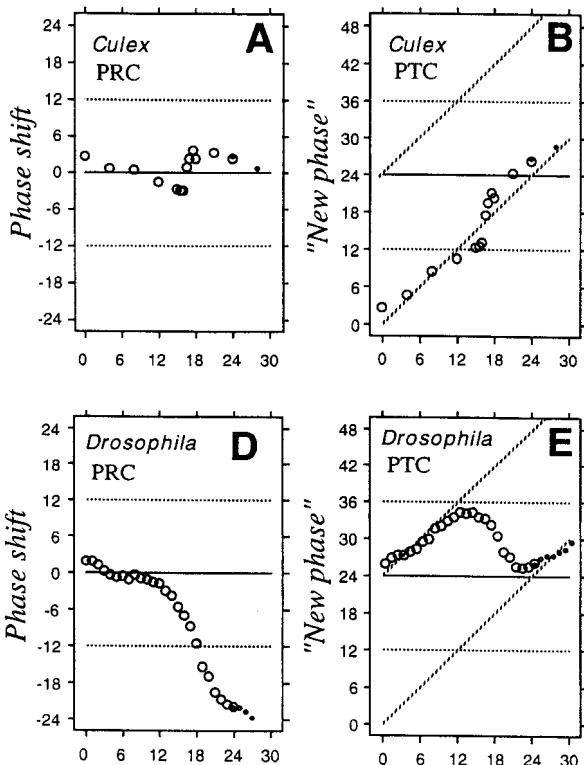


Figure 9. Resetting represented in both phase response curves (left) and phase transition curves (right). From Johnson (1999).

## Discussion

Of the representational formats used by scientists in performing and communicating their research, diagrams are particularly important but not yet extensively studied. We began our exploration of this topic by examining the use of diagrams to delineate phenomena. More specifically, we examined how actograms, phase response curves, and phase transition curves each highlight, and enable scientists to see, different circadian phenomena. Once delineated, such phenomena are explained by working out the responsible mechanism, and in later work we will examine other kinds of diagrams and their roles in mechanistic explanation.

We have drawn attention to the fact that scientists are often developing new representational formats to make manifest the specific phenomena in which they are interested. It is important to note that the researchers who devised the formats in Figures 1-9 were not seeking the best way of representing all circadian phenomena in a single diagram; rather, each sought to elucidate a specific phenomenon, such as changes in phase or the susceptibility to entrainment by light. In highlighting one phenomenon, each format either obscures or leaves out others. Actograms make clear how the phase of rhythms changes when an animal is switched to a free-running condition or exposed to a specific perturbation, but they do not show whether the rhythm is dampening or how the phase would change across the full range of possible times of perturbation. Phase response curves do this

latter job, but do not display phase changes after a switch to free-running.

New representational formats make new cognitive demands on audiences. If these formats were new to you, you also experienced the learning that is required to see what each diagrammatic format is representing. Only after learning to see the phenomenon in the diagram can the scientist use that format to efficiently reason about the phenomenon and begin the work of explaining it.

## References

Aschoff, J., & Wever, R. A. (1981). The circadian system of man. In J. Aschoff (Ed.), *Biological rhythms. Volume 4 of Handbook of Behavioral Neurobiology* (pp. 311-331). New York: Plenum.

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.

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

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

Hastings, J. W., & Sweeney, B. M. (1958). A persistent diurnal rhythm of luminescence in *Gonyaulax polyedra*. *Biological Bulletin*, 115, 440-458.

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

Hempel, C. G. (1965). Aspects of scientific explanation. In C. G. Hempel (Ed.), *Aspects of scientific explanation and other essays in the philosophy of science* (pp. 331-496). New York: Macmillan.

Johnson, C. H. (1999). Forty years of PRCs-What have we learned? *Chronobiology International*, 16, 711-743.

Johnson, M. S. (1926). Activity and distribution of certain wild mice in relation to biotic communities. *Journal of Mammalogy*, 7, 254-277.

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

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

Pittendrigh, C. S. (1960). Circadian rhythms and the circadian organization of living systems. *Cold Spring Harbor Symposia on Quantitative Biology*, 25, 159-184.

van Essen, D. C., & Gallant, J. L. (1994). Neural mechanisms of form and motion processing in the primate visual system. *Neuron*, 13, 1-10.