Diagrammatic Reasoning

William Bechtel

Summary

Diagrams figure prominently in human reasoning, especially in science. Cognitive science research has provided important insights into the inferences afforded by diagrams and revealed differences in the reasoning made possible by physically instantiated diagrams and merely imagined ones. In scientific practice, diagrams figures prominently both in the way scientists reason about data and in how they conceptualize explanatory mechanisms.

To identify patterns in data, scientists often graph it. While some graph formats, such as line graphs, are used widely, scientists often develop specialized formats designed to reveal specific types of patterns and not infrequently employ multiple formats to present the same data, a practice illustrated with graph formats developed in circadian biology. Cognitive scientists have revealed the spatial reasoning and iterative search processes scientists deploy in understanding graphs.

In developing explanations, scientists commonly diagram mechanisms they take to be responsible for a phenomenon, a practice again illustrated with diagrams of circadian mechanisms. Cognitive science research has revealed how reasoners mentally animate such diagrams to understand how a mechanism generates a phenomenon.

1. Overview

Human reasoning is often presented as a mental activity in which we apply inference rules to mentally represented sentences. In the 19th century Boole presented the rules for natural deduction in logic as formalizing the rules of thought. Even as cognitive scientists moved beyond rules of logical inference as characterizing the operations of the mind, they tended to retain the idea that cognitive operations apply to representations that are encoded in the mind (e.g., in neural activity). But in fact humans often reason by constructing, manipulating, and responding to external representations, and this applies as well to deductive as to abductive and inductive reasoning. Moreover, these representations are not limited to those of language but include diagrams. While reliance on diagrams extends far beyond science, it is particularly important in science. Scientific papers and talks are replete with diagrams and these are often the primary focus as scientists read papers and engage in further reasoning about them.[1] They also figure prominently in the processes

---

[1] Zacks, Levy, Tversky, and Schiano (2002) determined that the number of graphs in scientific journals doubled between 1984 and 1994. One would expect that trend has continued. Although many journals now limit the number of figures that can appear in the published paper, they have increasing allowed authors to post supplemental material, which often includes many additional diagrams. Scientists clearly use these diagrams to communicate their results with others. But there is also evidence that they make extensive use of these diagrams in their own thinking—developing an understanding of the phenomenon to be

through which scientists analyze data and construct their explanations. Since far less attention has been paid, both in philosophy of science and in the cognitive sciences, to how diagrams figure in reasoning activities, my objective in this chapter is to characterize what is known about how people, including scientists, reason with diagrams.

An important feature of diagrams is that they are processed by the visual system, which in primates is a very highly developed system for extracting and relating information received by the eyes (approximately one third of the cerebral cortex is employed in visual processing). I begin in section 2 by focusing on the distinctive potential of diagrams to support reasoning by enabling people to employ visual processing to detect specific patterns and organize together relevant pieces of information and examine the question of whether images constructed in one's imagination work equally well. In this chapter I employ the terms *diagram* in its inclusive sense in which it involves marks arranged in a two or more dimensional layout where the marks are intended to stand for entities or activities or information extracted from them and the geometrical relations between the marks are intended to convey relations between the things represented. In sections 3 and 4 I will discuss separately two types of diagrams that I designate data graphs and mechanism diagrams. In each case I introduce the discussion with examples from one field of biological research, that on circadian rhythms—the endogenously generate oscillations with a period of approximately 24 hours that are entrainable to the light-dark cycle of our planet and that regulate a wide range of physiological activities. I then draw upon cognitive science research relevant to understanding how people reason with each type of diagram and relate this to the diagrams used in the science.

2. Cognitive Affordances of Diagrams and Visual Images

Two different traditions have dominated cognitive science research on vision. One, associated with Marr (1982), has emphasized how, from the activation of individual neurons in the retina, people can build up a representation of what is seen. The other, advanced by Gibson (1979) drew attention to the rich information, often highly structured, available to the visual system. The latter is especially relevant to addressing diagrams, since they involve structured perceptual objects in the environment. A key theoretical claim Gibson advanced was that different objects of perception afford different activities for different organisms—the back of a chair affords landing for an insect but draping a garment for humans. One can extend the account of affordance to external representations, and so focus, as Zhang (1997) does, on how different representations activate different cognitive operations:

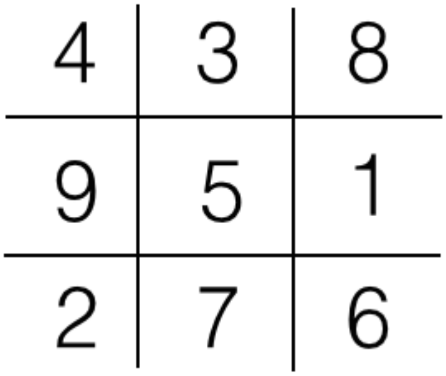
> Different representations activate different operations, not vice versa. It follows that operations are representation-specific. External representations activate perceptual operations, such as searching for objects that have a common shape and inspecting whether three objects lie on a straight line. In addition, external representations may have invariant information that can be directly perceived . . . such as whether several objects are spatially symmetrical to each other and whether one group has the same number of objects as another group. Internal representations activate cog-

---

explained and explanatory accounts. They do so in large part to enable the use of visual processing to identify patterns in the diagrams.

> nitive operations, such as adding several numbers to get the sum." (pp. 185-6)

To investigate how diagrams afford different reasoning that other representations, Zhang compared a game formally equivalent to tic-tac-toe in which players pick numbers from the pool 1 through 9 with the objective of being the first to pick three numbers totaling 15. Representing the numbers on a tic-tac-toe board (Figure 1) shows that the two games are formally equivalent—all sequences of three numbers totaling 15 can be mapped onto a winning solution to tic-tac-toe and vice versa. Despite being formally equivalent, the tic-tac-toe board representation engages different cognitive operations than the number game represented as picking numbers from a pool. On the tic-tac-toe board players can identify winning combinations by detecting lines but in the number variant they must perform arithmetic over many sets of numbers. In Zhang's experiments, humans played against a computer, which always made the first play and was programmed never to lose. If participants chose the best moves, however, they could always gain a tie. Participants required much longer to figure out a strategy to tie the computer when playing the number version than traditional tic-tac-toe, indicating that they deployed different operations in the two games. (See Zhang & Norman, 1994, for experiments showing similar results with variants of the Tower of Hanoi problem that placed different demands on internal processes.) Zhang further claims that by limiting winning strategies to lines, traditional tic-tac-toe reduces the cognitive demands, freeing up cognitive resources for other activities.



4	3	8
9	5	1
2	7	6

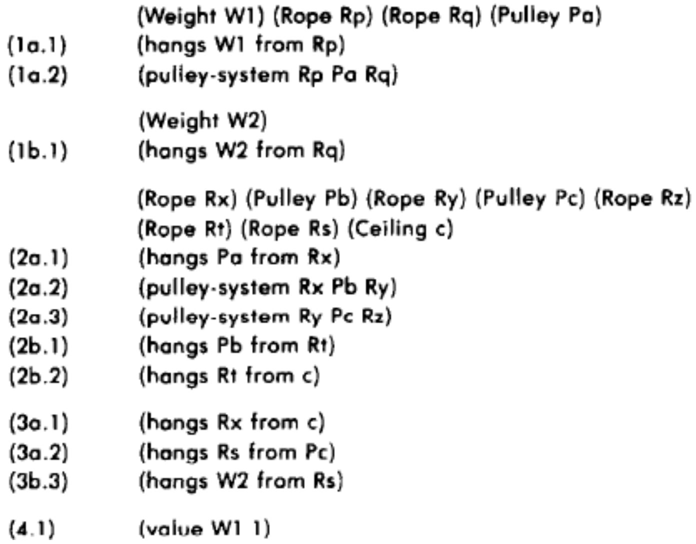
Figure 1. The game of picking three numbers that add to 15 is mapped onto a tic-tac-toe board, establishing their formal equivalence.

In a provocative pioneering study addressing the question "why a diagram is (sometimes) worth 10,000 words?" Larkin and Simon (1987) also focused on how diagrammatic representations support different cognitive operations appropriate than sentential representations. Like Zhang, they focused on representations that were equivalent in the information they provided but turned out not to be computationally equivalent in the sense that inferences that could be "drawn easily and quickly from the information given explicitly in the one" could not be drawn easily and quickly from the other.[2] One of the problems they in-

---

[2] Kulvicki (2013) speaks in terms of information being extractable where there is a feature of a representation that is responsible for it representing a given content and nothing more specific than that. This helpfully focuses on the issue of how the representation is structured, but does not draw out the equally important point that extracting information depends on the cognitive processes that the cognizer employs.

vestigate is the pulley problem shown in Figure 2, where the task is to find the ratio of weights at which the system is in equilibrium. They developed a set of rules to solve the problem. The advantage of the pulley diagram on their analysis is that it locates information needed to apply particular rules at nearby locations in the diagram so that by directing attention to a location a person can secure the needed information. In the sentential representation the information needed for applying rules was dispersed so that the reasoner would need to conduct multiple searches. In a second example, involving a geometry proof, Larkin and Simon show how a diagram reduces both the search and recognition demands, where recognition utilizes the resources of the visual system to retrieve information. The authors also offer three examples of diagrams used in economics and physics, graphs and vector diagrams, that employ not actual space but dimensions mapped to space and argue that they too provide the benefits in search and recognition.


(Weight W1) (Rope Rp) (Rope Rq) (Pulley Pa)
(1a.1) (hangs W1 from Rp)
(1a.2) (pulley-system Rp Pa Rq)

(Weight W2)
(1b.1) (hangs W2 from Rq)

(Rope Rx) (Pulley Pb) (Rope Ry) (Pulley Pc) (Rope Rz)
(Rope Rt) (Rope Rs) (Ceiling c)
(2a.1) (hangs Pa from Rx)
(2a.2) (pulley-system Rx Pb Ry)
(2a.3) (pulley-system Ry Pc Rz)
(2b.1) (hangs Pb from Rt)
(2b.2) (hangs Rt from c)

(3a.1) (hangs Rx from c)
(3a.2) (hangs Rs from Pc)
(3b.3) (hangs W2 from Rs)

(4.1) (value W1 1)

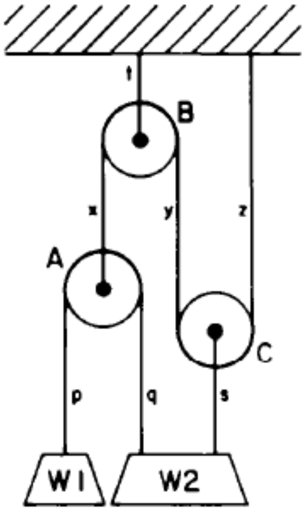


Figure 2. Larkin and Simon’s pulley problem presented in sentential form on the left and in a diagram on the right.

Together, these two studies make clear that diagrams differ from other representations in terms of the cognitive operations they elicit in problem solving situations. Most generally, diagrams as visual structures elicit pattern detection capacities whereas sentential representations require linguistic processing. Larkin and Simon note that a common response to a complex sentential description is to draw a diagram. An interesting question is whether comparable results can be obtained by mentally imagining diagrams. Pioneering studies by Shepard (Cooper & Shepard, 1973; Shepard & Metzler, 1971) and Kosslyn (Kosslyn, Ball, & Reiser, 1978) demonstrated that people can rotate or move their attention across a mentally encoded image. But quite surprisingly Chambers and Reisberg (1985) found that this capacity is severely limited. They presented Jastrow’s duck-rabbit (Figure 3) to participants sufficiently briefly that they could only form one interpretation of the figure. They then asked the participants if they could find another interpretation while imaging the figure. None were able to do so even when offered guidance. Yet, when they were allowed to draw a figure based on their mental image, all participants readily discovered the alternative interpretation.

Figure 3. The version of the duck-rabbit figure used as a stimulus in Chambers and Reisberg's experiments is shown on the left. The other two versions were drawn by participants based on their own image interpreted as a rabbit (center) and as a duck (right). From these they were readily able to discover the other interpretation, something they could not do from their mental image alone.

These findings inspired numerous other investigations into the human ability to work with mental images whose results present a complex pattern. Reed and Johnsen (1975) reached a similar conclusion as Chambers and Reisberg when they asked participants to employ imagery to determine whether a figure was contained in a figure they had previously studied. Yet when Finke, Pinker, and Farah (1989) asked participants to construct in imagery complex images from components, they performed well. Studies by Finke and Slayton (1988) showed that many participants were able to generate creative images from simple shapes in imagery (the drawings the participants produced were independently assessed for creativity). Anderson and Helstrup (1993a, 1993b) set out to explore whether drawing enhanced performance on such tasks and their conclusions were largely negative—participants produced more images, but the probability of generating ones judged creative was not increased: "These results were contrary to the initial belief, shared by most experimenters and subjects alike, that the use of pencil and paper to construct patterns should facilitate performance."

Verstijnen, Van Leeuwen, Hamel, and Hennessey (2000) explored whether the failure of drawing to improve performance might be due to insufficient training in drawing. Using a task similar to that of Reed and Johnson, they compared those without formal training in drawing with design students who had two years of courses in drawing, and found those with training in drawing performed much better. In another study in which participants were required to create new objects from simple components, Verstijnen, van Leeuwen, Goldschmidt, Hamel, and Hennessey (1998) found that drawing significantly helped trained drawers create compound objects that involved restructuring the components (e.g., changing proportions within the component). One conclusion suggested by these results is that reasoning with diagrams may be a learned activity. Humans spend a great deal of time learning to read and write, and even then further education is often required to extract information from text and construct and evaluate linguistic arguments. Yet, perhaps because vision seems so natural, we assume that diagrams are automatically interpretable and except in curricula in fields like design, we provide no systematic education in constructing and reasoning with diagrams. Accordingly, it perhaps should not be a surprise that science educators have found that students often ignore the diagrams in their textbooks (Cook, 2008). One of the challenges in teaching students how to reasoning with diagrams is identifying what cognitive operations people must perform with different types of diagrams.

Cognitive scientists have begun to identify some of these operations, and I will discuss some of these in the context of data graphs and mechanism diagrams in the next two sections.

3. Reasoning with Data Graphs

3.1 Data Graphs in Circadian Biology

By far the majority of the diagrams that figure in scientific papers are devoted to graphing data. Surprisingly, given the recognition of the roles data play both in discovering possible explanations and in evaluating them, there is little discussion in philosophy of graphing practices and how they figure in discovery and justification. Rather, the focus has been on data claims that can be represented sententially. Although there are common graphic formats that are highly familiar—e.g., line graphs and bar graphs—in fact a wide variety of graphic formats are frequently used in science. In particular fields scientists have created their own formats, but these formats often migrate between fields. Each format elicits specific visual processing operations to identify informative patterns. In addition to different graphic formats, there are different tasks in which scientists present data. I focus on two tasks—delineating phenomena and presenting relations between variables that are taken to be explanatory of the phenomenon.

In presenting phenomena as the target of scientific explanations, Bogen and Woodward (1988) distinguish phenomena from data. They argue that phenomena, unlike data, are repeatable regularities in the world. Data provide evidence for the occurrence of phenomena. In many cases, researchers delineate phenomena by identifying patterns in data they collect. In the case of circadian rhythms, these are patterns of activity that repeat every 24 hours and at detectable by visual inspection of diagrams.

One of the most basic diagramming techniques employs a Cartesian coordinate system on which one plots values of relevant variables on the two axes. Using the abscissa to represent time and the ordinate for the value of a variable such as temperature, circadian researchers can plot each data point and then connect them by lines or a smoothed curve (Figure 4, left). Our visual system readily identifies the oscillatory pattern, which we can then coordinate with the bar at the bottom indicting periods of light and dark and the grey regions that redundantly indicate periods of darkness. By visually investigating the graph, one can detect that body temperature rises during the day and drops during the night, varying by about 2° F. over the course of a day.

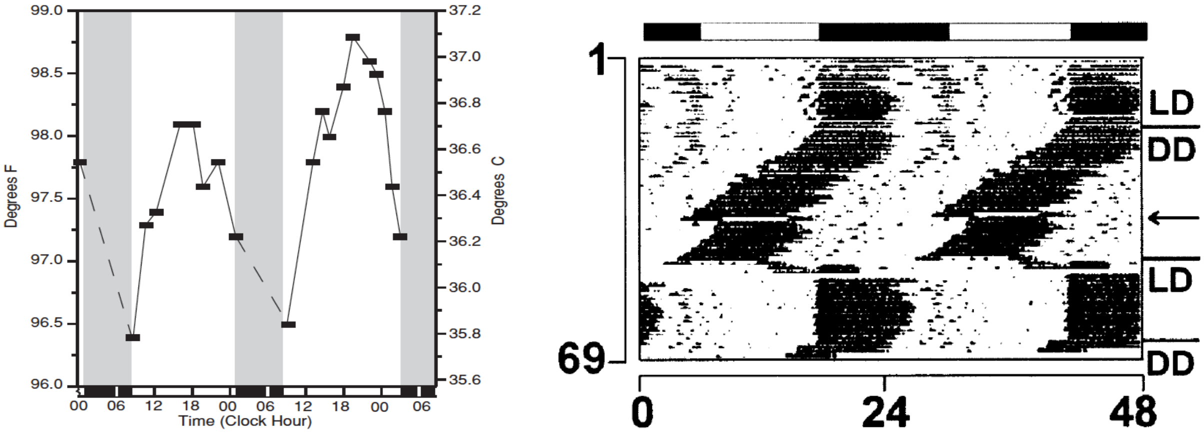


Figure 4. Left: Line graph from Koukkari and Southern (2006) showing the circadian oscillation in body temperature for one person across 48 hours. Right: Example actogram showing times of running wheel activity of a wild-type mouse from Bunger, Wilsbacher, Moran, Clendenin, Radcliffe, Hogenesch, Simon, Takahashi, and Bradfield (2000).

A line graph makes clear that the value of a variable is oscillating and with what amplitude, but it does not make obvious small changes in the period of activity. For this reason, circadian researchers developed actograms—a version of a raster plot on which time of each day is represented along a horizontal line and each occurrence of an activity (rotation of a running wheel by a mouse) is registered as a hash mark. Subsequent days are shown on successive lines placed below the previous one. Some actograms, such as the one shown on the right in Figure 4, double plot the data so that each successive 24-hour period is both plotted to the right of the previous 24-hour period and then again on the left on the next line. Placing adjacent times next to each other even when they wrap around a day break makes it easier to track detect continuous activity patterns. An actogram renders visually apparent how the phase of activity changes under different conditions such as exposure to light. In this actogram, the mouse was first exposed to a 12:12 light dark cycle, as indicated by the letters LD on the right side, with the periods of light and dark indicated by the light-dark bar at the top. From day 15 to day 47, as indicated by the letters DD on the right side, the mouse was subjected to continuous darkness. On day 37, the row indicated by the arrow, the animal received a six-hour pulse of light at hour 16. It was returned to LD conditions on day 48, but returned to DD on day 67. The activity records shown on the actogram exhibit a clear pattern. During both LD periods the activity of the mouse was entrained to the pattern of light and dark so that the mouse was primarily active during the early night, with a late bout of activity late in the night (mice are nocturnal animals). On the other hand, during the DD periods the mouse began its activity somewhat earlier each day, a phenomenon known as *free running*. The light pulse reset the onset time for activity on the following day, after which the mouse continued to free run but from this new starting point. When switched back to LD the mouse exhibited a major alternation in activity the next day, but it took a couple more days to fully re-entrain to the LD pattern.

Data graphs are used not just to characterize phenomena but also to identify factors that may play a role in explaining phenomena. Figures in biological papers often contain many panels, invoking different representational formats, as part of the attempt to make visible relationship between variables that are taken to be potentially explanatory. For example, Figure 5, from Maywood, Reddy, Wong, O'Neill, O'Brien, McMahon, Harmar, Okamura, and Hastings (2006), employs photographs, line graphs, heat maps, and radial (Rayleigh) plots. To situate their research, in the 1970s the suprachiasmatic nucleus (SCN), a small structure in the hypothalamus, was implicated through a variety of techniques as the locus of circadian rhythms in mammals. Welsh, Logothetis, Meister, and Reppert (1995) had demonstrated that while individual SCN neurons maintain rhythmicity when dispersed in culture, they oscillate with varying periods and quickly become desynchronized. Maywood et al.'s research targeted vasoactive intestinal polypeptide (VIP), which is released by some SCN neurons, as the agent that maintains synchrony in the whole SCN or in slices from the SCN. Accordingly, they compared SCN slices from mice in which one (identified as VIP2r$^{+/-}$) or both copies (VIP2r$^{-/-}$) of the gene that codes for the VIP receptor are deleted. To render the rhythmicity of individual cells visible, the researchers inserted a gene coding for luciferase under control of the promoter for a known clock gene, *Per1*, so as to produce luminescence whenever PER is synthesized. The photographs in panels A are selections among the raw data. They make clear that VIP luminescence in the slice is synchronized, occurring at hour 48 and 72. Panel E reveals the lack of synchrony without the VPN receptor and panel F demonstrates that individual neurons are still oscillating without VIP but that the three neurons indicted by green, blue, and red arrows exhibit luminescence at different phases.

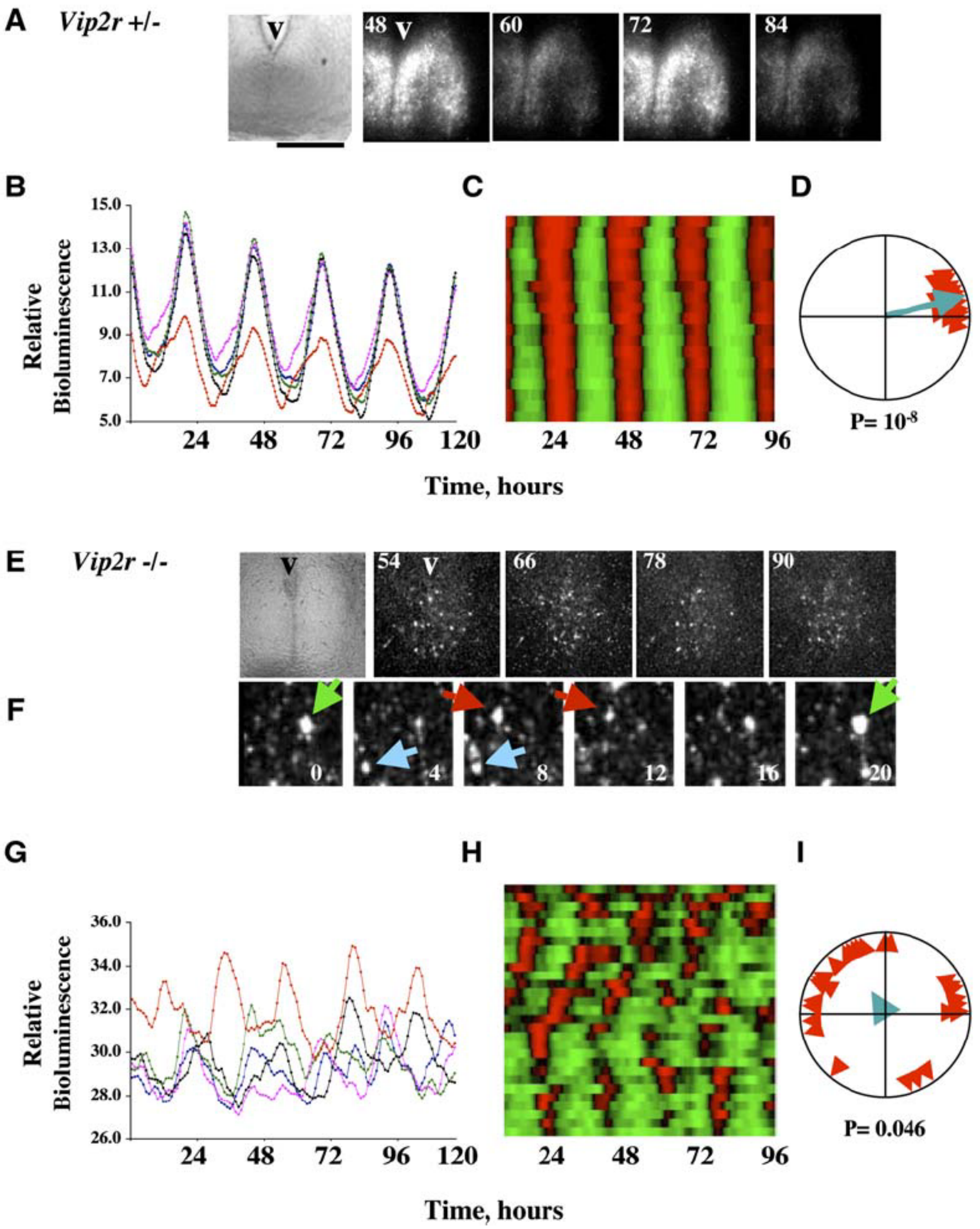


Figure 5. By using multiple graphical formats (photographs of slices from the SCN, line graphs, raster plots (heat maps) and radial (Rayleigh) plots, Maywood, Reddy, Wong, O'Neill, O'Brien, McMahon, Harmar, Okamura, and Hastings (2006) make apparent, when a receptor for VIP is present, oscillations of individual neurons are synchronized but that this is lost without VIP.

Although the photographs are sufficient to show that VIP is potentially explanatory of synchronous activity in the SCN, the researchers desired to characterize the relationship in more detail. They began by quantifying the bioluminescence recorded at the locus of the cell in photographs at different times. In panels B and G they displayed the results for five individual cells in each of each type in line graphs. This makes it clear that while there is variation in amplitude, with VIP the five cells are in phase with each other while without

VIP they are not. Even with five cells, though, it becomes difficult to decipher the pattern in a line graph. The raster plots in panels C and H enable comparison of 25 cells, one on each line, with red indicating periods when bioluminescence exceeds a threshold and green periods when it is below the threshold (such displays using hot and cold colors are often called *heat maps*). The raster plot enables one to compare the periodicity of individual cells more clearly, but with a loss of information about the amplitude of the oscillation at different times. The Rayleigh plots shown in panels D and I sacrifice even more information, focusing only on peak activity, but show that the peak phases are highly clustered with VIP and widely distributed without. The blue arrow shows the aggregate phase vector and indicates not only that it is oriented differently without VIP but also is extremely short, indicative of little correlation between individual neurons.

3.2 Cognitive science research relevant to reasoning with graphs

Having introduced examples of graphs used in one field of biology, I turn now to cognitive science research that has attempted to identify aspects of the cognitive operations that figure in reasoning with graphs. Pinker (1990) provided the foundation for much subsequent research on how people comprehend graphs. He differentiated the cognitive activities of creating a visual description of a graph and applying an appropriate graph schema to it. He treats the construction of a visual description as initially a bottom up activity driven by the visual stimulus to which gestalt principles such as proximity and good continuation, among other procedures, are invoked. As explored by Zacks and Tversky (1999), these principles differentially affect perception of bar graphs and line graphs: “Bars are like containers or fences, which enclose one set of entities and separate them from others. Lines are like paths or outstretched hands, which connect separate entities” (p. 1073). The result, which has been documented in many studies, is that people are faster and more accurate at reading individual data points from bar graphs than line graphs but detect trends more easily in line graphs (Simcox, 1984; Carswell & Wickens, 1987). For example, the bar graph on the left in Figure 6 makes it easy to read off test scores at different noise levels and room temperatures, and to compare test scores at the two temperatures. The line graph in the middle encodes the same data but the lines connecting the values at the two noise levels make that comparison more apparent. Moreover, the line graph suggests that there are intermediate values between the two explicitly plotted. The effect is sufficiently strong that Zacks and Tversky found that when line graphs are used with categorical variables, viewers often treat them as interval variables and make assertions such as “The more male a person is, the taller he/she is” (p. 1076).

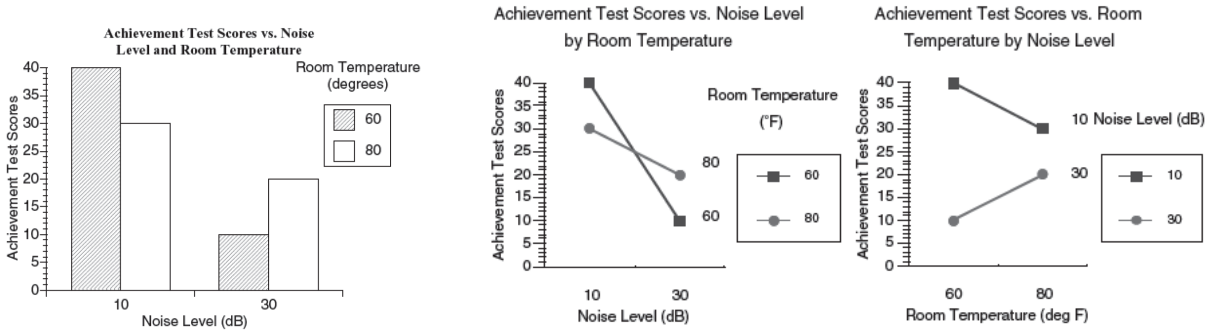


Figure 6. A bar graph on the left and two line graphs, each showing the same data, but which viewers typically interpret differently. From Shah and Hoeffner (2002).

The choice of what to present on the axes also affects the information people extract. Shah and Carpenter (1995) found that participants produce very different interpretations of the two graphs on the right of Figure 6, one representing noise and the other room temperature on the abscissa. Thus, viewers of the graph in the center are more likely to notice the trend with increasing noise levels whereas those viewing the graph on the right notice the trend with increasing temperature. Further, when lines in graphs have reverse slopes, as in the rightmost graph, participants take longer to process the graph. Moreover, this difference makes the third variable, noise level, more salient since it identifies the difference responsible for the contrasting slopes.

The research reported so far focused on visual features of graphs, but one of the seminal findings about the organization of the mammalian visual processing system is that it is differentiated into two processing streams, one extracting information about the shape and identity of objects and one extracting information about location and potential for action (Ungerleider & Mishkin, 1982; van Essen & Gallant, 1994). Hegarty and Kozhevnikov (1999) proposed that the distinction between different processing pathways could help explain apparently contradictory results other researchers had reached about whether skill in visual imagery facilitates solving mathematics problems. They separately evaluated sixth-grade boys in Dublin, Ireland in terms of pictorial imagery ("constructing vivid and detailed images") and schematic imagery ("representing the spatial relationships between objects and imagining spatial transformations"). The found that good pictorial imagery was actually associated with poorer performance in solving mathematical problems,[3] while good spatial imagery was associated with better performance. In subsequent work Hegarty and her collaborators focused on kinetic problems involving graphs of motion and demonstrated a similar effect of pictorial versus spatial visualization. Kozhevnikov, Hegarty, and Mayer (2002) presented graphs such as that on the left in Figure 8 to participants who, on a variety of psychometric tests, scored high or low on spatial ability. Those who scored low interpreted this graph pictorially as, for example, a car moving on a level surface, then going down a hill, and then moving again along a level surface. None of these participants could provide the correct interpretation of the graph as showing an object initially at rest, then moving at a constant velocity, and finally again at rest. On the other hand, all participants who scored high on spatial ability provided the correct interpretation. Subsequently, Kozhevnikov, Kosslyn, and Shephard (2005) examined the differences between professionals in the arts and the sciences with respect to these graphs. They found that, except for participants who provided an irrelevant interpretation by focusing on non-pictorial features of the graph, artists tended to provide a literal pictorial interpretation of the path of movement whereas all scientists offered a correct schematic interpretation (example responses are shown on the right of figure 7).

---

[3] The following was a typical problem: "At each of the two ends of a straight path, a man planted a tree and then every 5 meters along the path he planted another tree. The length of the path is 15 meters. How many trees were planted?"

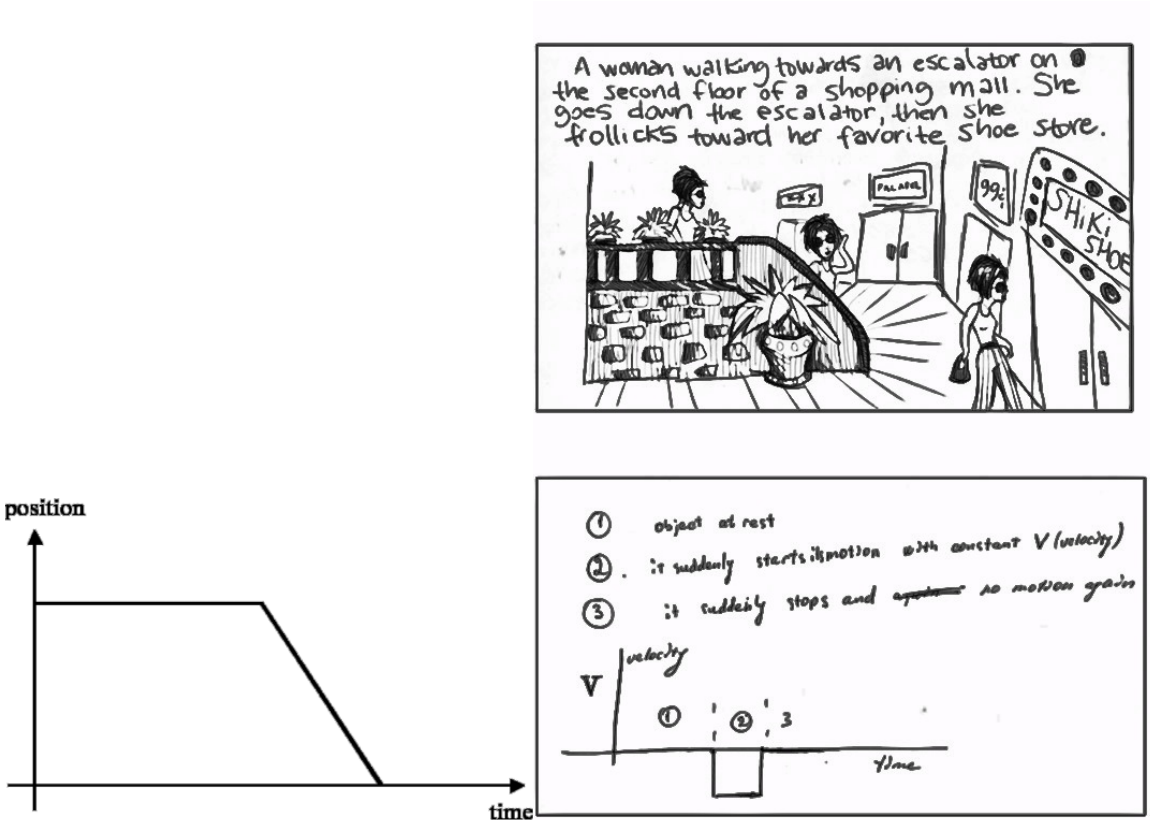


Figure 7. On the left is a line graph showing an object initially at rest, then moving for a period at a constant velocity, then returning to rest that Kozhevnikov, Kosslyn, and Shephard (2005) used to compare interpretations by artists and scientists. A typical response from an artist is shown at the top right whereas one from a scientist is shown at the bottom.

So far I have focused on viewing a graph and extracting information from it. But an important feature of graphs in science such as those I presented in the earlier section is that they afford multiple engagements in which a user visually scans different parts of the graph seeking answers to different questions, some posed by information just encountered. Carpenter and Shah (1998) drew attention to this by observing that graph comprehension is an extended activity often requiring half a minute, two orders of magnitude longer than the time required to recognize simple patterns, including words and objects. In addition to detecting a pattern of data points along, e.g., a positively sloping line, the graph interpreter must relate these points to the labels on the axes and what these represent and this is what requires processing time. Using eye tracking which participants study graphs, Carpenter and Shah revealed that viewers initially carve the graph into visual chunks and then cycle through focusing on different components—the pattern of the lines, the labels on the axes, the legend, and the title of the graph (Figure 8). Similarly drawing attention to the prolonged engagement individuals often have with graphs, Trickett and Trafton (2006) employed verbal protocols as well as eye tracking to study what people do when making inferences that go beyond what is explicitly represented in a given graph. They found that participants often employ spatial manipulations such as mentally transforming an object or extending it; they are not just passively viewing it.

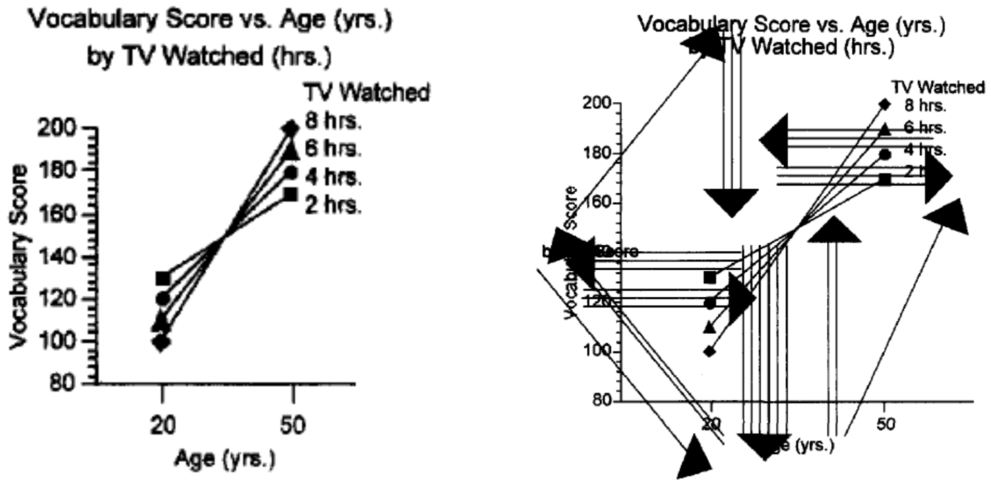


Figure 8. Graph (left) and superimposed eye-tracking results (right) from Carpenter and Shah (1998).

Cognitive scientists have limited their focus to relatively simple graph forms such as line graphs and have not investigated the larger range of format we saw deployed in circadian research. Many of the results, however, are applicable to these other graph formats. Gestalt principles such as good continuation affect the patterns people see in actograms and raster plots (heat maps). In the actogram in Figure 4 one recognizes the phase locking of activity to the light-dark cycle and daily phase advance when light cues are removed by implicitly (and sometimes explicitly) drawing a line through the starting point for each day's activity. Spatial processing is clearly important not only with the photographs in Figure 5 but also with the heat map and Rayleigh plot. A skilled user of these graphs must recognize that space in the photographs corresponds to space on the slice from the brain but that space in the heat map corresponds not to physical space but an abstract space in which different cells are aligned. Finally, these diagrams are not designed to convey information in one look but rather are objects that afford shifting one's attention many times to focus on different information. With the Rayleigh plot, for example, one typically attends separately to the dispersal of blue arrowheads reflecting peaks of individual cells and to the vector indicating the population average. If eye tracking were performed, the pattern would likely resemble that displayed in Figure 8. With panels showing the same information in multiple formats, as in Figure 5 viewers are also likely to shift their focus between panels to see, for example, how the times in the line graph correspond to those in the photograph or those in the heat map. One limitation of the cognitive science studies is that the tasks participants were asked to perform were usually quite limited (e.g., interpret the graph) whereas scientists often use interact with graphs over multiple engagements, constructing new queries on the basis of previous ones (e.g., probing an actogram to see if the behavior really does look rhythmic or not or exploring the variability between cells revealed in a heat map). This is particularly evident when a researcher pours over a graph after producing it to determine what it means or when, in a journal club discussion, other researchers raise questions about specific features of a graph. Ultimately we need to better understand how scientists pose and address such queries over time if we are to understand the different roles graphs play in scientific reasoning.

4. Reasoning with Mechanism Diagrams

4.1 Mechanism diagrams in circadian biology

Recognizing that individual activities, even if they do play a causal role in generating a phenomenon, typically do not work in isolation but only in the context of a mechanism in which they interact with other components, biologists often set as their goal to characterize the mechanism.[4] The researchers' conception of the mechanism is sometimes presented in a final figure in a journal article but mechanism diagrams are even more common in review papers. Figure 9 is a representative sample of a mechanism diagram for the intracellular oscillator in mammalian SCN neurons. The diagram uses glyphs—"simple figures like points, lines, blobs, and arrows, which derive their meaning from geometric or gestalt properties in context" (Tversky, 2011) to represent the parts and operations of the mechanism.[5] The parts shown include DNA strands, shown as two wavy lines, on which promoter regions are indicated by lightly-shaded rectangles, genes by darkly colored rectangles, and protein products, by colored ovals. Lines with arrow heads represent operations such as expression of a gene or transport of proteins to locations where they figure in other reactions, including activating gene transcription. Lines with squared ends indicate inhibitory activity. When phosphates attach to molecules (as preparation for nuclear transport or degradation), they are are shown as white circles containing a P.

---

[4] For discussion of the appeals to mechanism to explain biological phenomena, see Bechtel and Richardson (1993/2010); Bechtel and Abrahamsen (2005); Machamer, Darden, and Craver (2000).

[5] Tversky emphasizes the abstractness of glyphs over more iconic representations, arguing that the abstractness promotes generalization. One can abstract even more by allowing only one type of glyph (e.g., a circle) for a entity and one for an operation (an arrow), generating the sort of representations found in graph theory and used to capture general consequences of the organization of mechanisms. See Bechtel (in press).

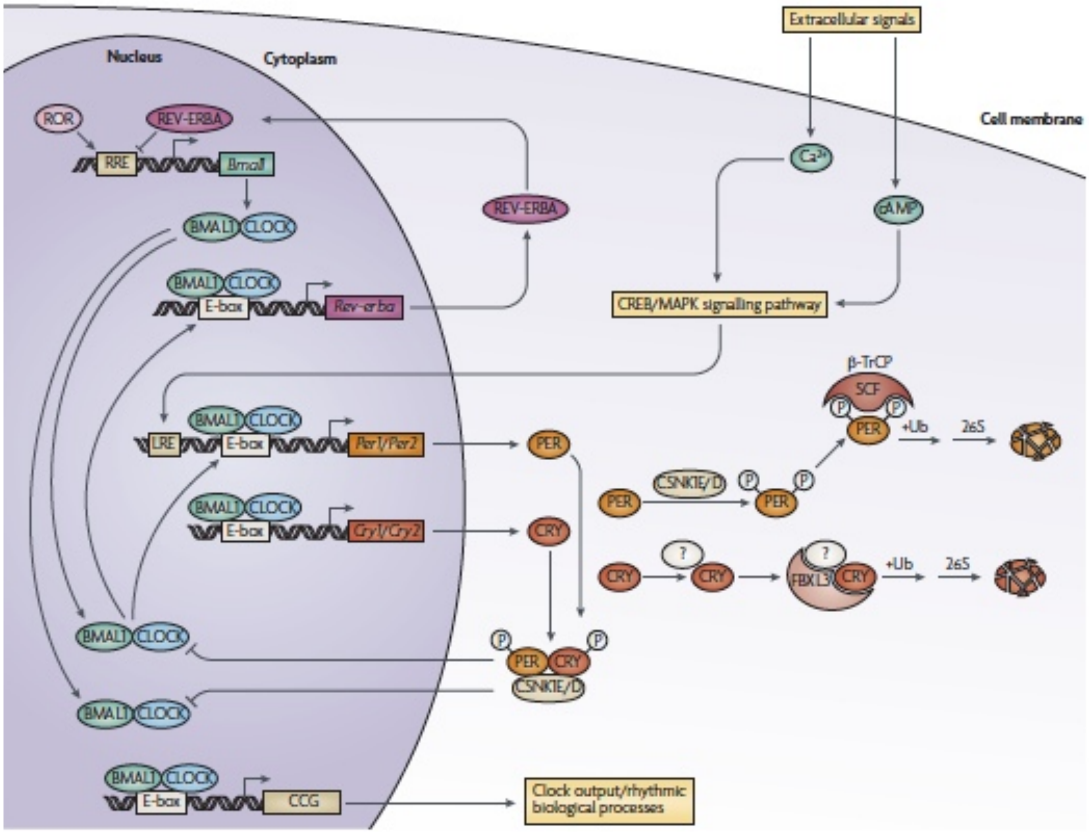


Figure 9. Takahashi, Hong, Ko, and McDearmon (2008) mechanism diagram of the mammalian circadian clock involve genes and proteins within individual cells.

The diagram is clearly laid out spatially, but only some features of the diagram convey information about spatial structures in the cell. The differentiation of the nucleus and cytoplasm is intended to correspond to these regions in the cell and lines crossing the boundary between the nucleus and cytoplasm represent transport between the two parts of the cell. Beyond that, however, the distribution of shapes and arrows conveys no spatial information but only functional differentiation. The most important operations shown in this diagram are the synthesis of REV-ERBA and its subsequent transport into the nucleus to inhibit transcription of BMAL1 (shown as a loop out from and back into the nucleus in the upper left) and the synthesis of PER and CRY, the formation of a dimer, and the transport of the dimer into the nucleus to inhibit the ability of BMAL1 and CLOCK to activate transcription of BMAL1, PER, and CRY (shown as a loop out from and back into the nucleus in the center-left of the figure). (The other operations shown are those involved in signaling from outside the cell that regulates the overall process, in the degradation of PER and CRY, and in the expression of Clock-Controlled Genes [CCG] that constitute the output of the clock.

For someone acquainted with the types of parts shown and the operations in which they engage, a diagram such as this provides a means of showing schematically how the various parts perform operations that affect other parts. One is not intended to take in the whole diagram at once, but to follow the operations from one part to another. To understand how the mechanism gives rise to oscillatory activity, one can mentally simulate the operations of the mechanism by starting in the middle with the *Per* and *Cry* genes. As they are expressed, more PER and CRY proteins are generated. After the proteins form a dimer and are transported into the nucleus, they inhibit the activity of the BMAL1:CLOCK dimer and thereby stop their own expression. This reduction in express results in reduction in their concentration and reduced inhibitory activity, which allow expression to resume. This capacity for mental animation is, however, limited, and to determine what the activity will be, especially when the other components are included, researchers often turn to computational models, generating what Abrahamsen and I (Bechtel & Abrahamsen, 2010) refer to as *dynamic mechanistic explanations*. Even here, though, diagrams provide a reference point in the construction of equations describing operation of the various parts (Jones & Wolkenhauer, 2012).

Looking carefully at the lower right side of the figure, one will see two ovals with question marks in them. This indicates that the researchers suspected that something unknown binds with CRY before and potentially mediates its binding with FBXL3, which then results in its degradation. Here it is the identity of an entity that is in doubt, but sometimes question marks are employed to indicate uncertainty about the identity of an operation. In this case the diagram is from a review paper and the question mark reflects uncertainty in the discipline. On occasions when question marks appear in mechanism diagrams presented at the beginning of a researcher paper they signal that the goal of the paper is to answer a question regarding the identity of a component or its operation.

4.2 Cognitive science research relevant to reasoning with mechanism diagrams

Although cognitive scientists have not explicitly focused on mechanism diagrams that figure in biology,[6] research on simple mechanical systems such as pulley systems (already the focus of Larkin and Simon's research discussed above) has highlighted one of the important cognitive activities people use with mechanism diagrams—mentally animating the operation of a mechanism when trying to figure out how it will behave. Drawing on theorists in the mental models tradition (see papers in Gentner & Stevens, 1983, that explore how people answer problems by constructing and running a mental model), Hegarty (1992) investigated experimentally "to what extent the mental processes involved in reasoning about a mechanical system are isomorphic to the physical processes in the operation of the system." She measured reaction times and eye movements as participants answered questions about how various parts of a pulley systems such as shown in Figure 10 would behave if the rope is pulled. From the fact that both error rates and reaction times increased with the number of operations within the mechanism the participant had to animate in order to answer the question, she inferred that people don't simulate the whole machine operating at once but rather animate individual parts in sequence. She provided further evidence for this claim by tracking the movements of the participants' eyes as they solved problems. In a follow-up experiment, Hegarty compared performance when participants were asked to infer the motion of a component from that of another component earlier in the causal chain or from that of a component later in the chain. Participants made more errors and required more time when they had to reason backwards from events later in the chain, and still showed a preference to move their eyes forward along the causal chain.

---

[6] Stieff, Hegarty, and Dixon (2010) have explored strategies used to transform diagrams of molecular structure used in organic chemistry.

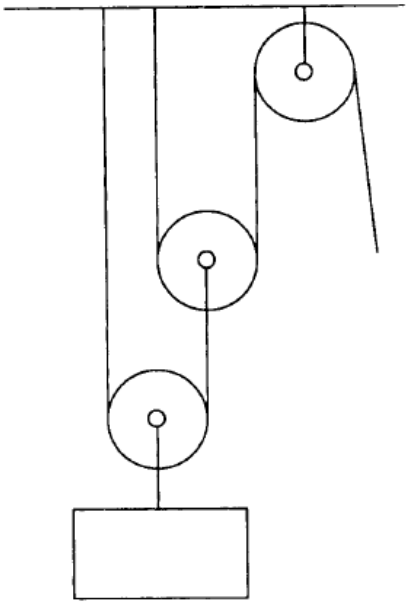
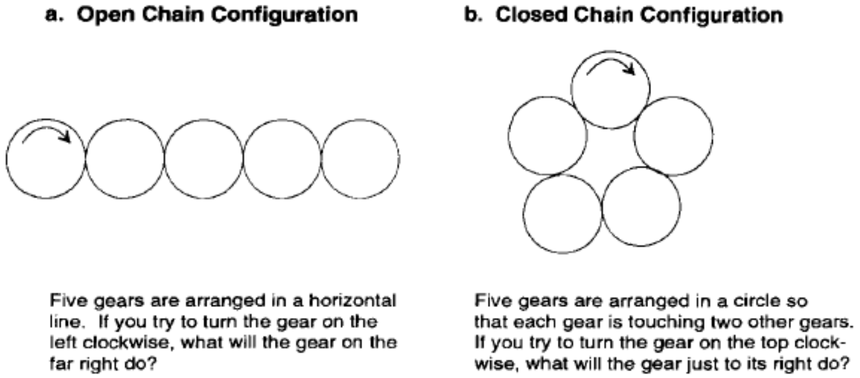


Figure 10. Pulley problem Hegarty used to study how people employ mental animation in problem solving.

Schwartz and Black (1996) provided further insights into how people simulate mechanisms by attending to the gestures people make. In one task, shown on the left below, participants were asked to determine in which direction the rightmost gear would turn given the clockwise turn of the leftmost gear. They found that their participants would use their hands to indicate the direction of movement of each successive gear. (In these studies the participants never saw the diagrams but were provided with verbal descriptions of the configuration.) In this case, an alternative strategy is available: apply a simple global rule such as the parity rule: "if there are an odd number of gears, the first and last will turn in the same direction" or the more local rule "if two gears are touching, they will turn in opposite directions." Schwartz and Black found that as people acquired the rule, their gestures declined. But when people lack such rules or find their application uncertain, as in the gear problem on the right in Figure 11, they again gesture. This use of gesture indicates that whatever imagery people employ to solve the task, it is coordinated with action. Accordingly, the researchers propose a theory of simulated doing in which "the representation of physical causality is fundamental. This is because 'doing' requires taking advantage of causal forces and constraints to manipulate the world. Our assumption is that people need to have representations of how their embodied ideas will cause physical changes if they are to achieve their goals" (Schwartz & Black, 1999).



a. Open Chain Configuration

Five gears are arranged in a horizontal line. If you try to turn the gear on the left clockwise, what will the gear on the far right do?

b. Closed Chain Configuration

Five gears are arranged in a circle so that each gear is touching two other gears. If you try to turn the gear on the top clockwise, what will the gear just to its right do?

Figure 11. Gear problem used by Schwartz and Black (1996) to study when participants gesture while solving problems.

Animating a diagram, either mentally or with gesture, plays a crucial role in the cognitive activity of understanding how a proposed mechanism could produce the phenomenon one is trying to explain. But diagrams present not only a finished explanation of the phenomenon, they often figure in the process of discovering mechanisms. Here what matters is the ability to create and alter the glyphs and their arrangement. Tversky (2011) suggests a helpful way to understand this activity—view diagrams as the "permanent traces of gestures" in which "fleeting positions become places and fleeting actions become marks and forms" (p. 500). There is a rich literature showing how gesture figures not only in communication but also in the development of one's own understanding (Goldin-Meadow & Wagner, 2005). Tversky focuses on the activity of drawing maps, highlighting such features of the activity as selecting what features to include and idealizing angles to right angles. These findings can be extended to mechanism diagrams, which constitute a map of the functional space of the mechanism, situating its parts and operations. While Tversky speaks of diagrams as permanent traces and there is a kind of permanence (or at least endurance) to diagrams produced on paper or in computer files, they are also subject to revision—one can add glyphs for additional parts or alter arrows to represent different ideas of how the operations of one part affect others. In the design literature this is often referred to as *sketching*. Sketching mechanism diagrams can be motivated by evidence, but they can also be done in a purely exploratory manner, enabling reasoning about what would happen if a new connection were made or an existing one redirected. Sketching possible mechanisms is a common activity of scientists, and by further investigating the cognitive activities involved in this activity one can develop richer analyses of this important type of scientific reasoning.

5. Conclusions and Future Tasks

This chapter has addressed the use of diagrams by scientists in characterizing phenomena to be explained, identifying variables that figure in explaining those phenomena, and advancing proposals for mechanisms, drawing examples from circadian rhythm research. Over the last thirty years cognitive scientists have attempted to characterize cognitive activities people employ when perceiving and using diagrams in problem solving tasks, such as making multiple scans of graphs and animating mechanical diagrams. For the most part, cognitive scientists have employed diagrams and tasks in their studies that are simpler than those that figure in actual scientific research. But these cognitive science studies nonetheless provide insights into the cognitive processes that figure in scientists' use of diagrams. To date the roles diagrams play in science has not figured in a major way in philosophical accounts of scientific reasoning but given the important roles diagrams play in science, there is great potential to advance our understanding of scientific reasoning by investigating further the cognitive processes involved as scientists create and use diagrams in the course of their research.

Acknowledgements

Research for this chapter was supported by Grant 1127640 from the U.S. National Science Foundation and a Senior Visiting Fellowship in the Center for Philosophy of Science at the University of Pittsburgh, which are gratefully acknowledged. I also thank fellow members of the Working Group On Diagrams in Science at UCSD (Adele Abrahamsen, Daniel C. Burnston, and Benjamin Sheredos) and John Norton and the Fellows at the Center for Philosophy of Science at the University of Pittsburgh in Fall 2014 (Joshua Alexander, Karim Bschir, Ingo Brigandt, Sara Green, Nicholaos Jones, Raphael Scholl, and Maria Serban) for insights and suggestions.

References

Anderson, R. E., & Helstrup, T. (1993a). Multiple perspectives on discovery and creativity in mind and on paper. In B. Roskos-Ewoldsen, M. J. Intons-Peterson & E. A. Rita (Eds.), *Imagery, creativity, and discovery: A cognitive perspective* (Vol. Volume 98, pp. 223-253). Dordrecht: Elsevier.

Anderson, R. E., & Helstrup, T. (1993b). Visual discovery in mind and on paper. *Memory & cognition, 21*, 283-293.

Bechtel, W. (in press). Generalizing mechanistic explanations through graph-theoretic perspectives. In P.-A. Braillard & C. Malaterre (Eds.), *Explanation in Biology. An Enquiry into the Diversity of Explanatory Patterns in the Life Sciences*. Dordrecht: Spinger.

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

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

Bechtel, W., & 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.

Bunger, M. K., Wilsbacher, L. D., Moran, S. M., Clendenin, C., Radcliffe, L. A., Hogenesch, J. B., Simon, M. C., Takahashi, J. S., & Bradfield, C. A. (2000). Mop3 Is an essential component of the master circadian pacemaker in mammals. *Cell, 103*, 1009-1017.

Carpenter, P. A., & Shah, P. (1998). A model of the perceptual and conceptual processes in graph comprehension. *Journal of Experimental Psychology-Applied, 4*, 75-100.

Carswell, C. M., & Wickens, C. D. (1987). Information integration and the object display: An interaction of task demands and display superiority. *Ergonomics, 30*, 511-527.

Chambers, D., & Reisberg, D. (1985). Can mental images be ambiguous? *Journal of Experimental Psychology: Human Perception and Performance, 11*, 317-328.

Cook, M. P. (2008). Students' comprehension of science concepts depicted in textbook illustrations. *Electronic Journal of Science Education, 12*.

Cooper, L. A., & Shepard, R. N. (1973). Chronometric sudies of the rotation of mental images. In W. G. Chase (Ed.), *Visual Information Processing Symposium on Cognitiion, 8th Carnegie Mellon University* (pp. 75-175). New York: Academic Press.

Finke, R. A., Pinker, S., & Farah, M. J. (1989). Reinterpreting visual patterns in mental imagery. *Cognitive Science, 13*, 51-78.

Finke, R. A., & Slayton, K. (1988). Explorations of creative visual synthesis in mental imagery. *Memory & cognition, 16*, 252-257.

Gentner, D., & Stevens, A. L. (1983). *Mental models*. Hillsdale, N.J.: Erlbaum.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston: Houghton Mifflin.

Goldin-Meadow, S., & Wagner, S. M. (2005). How our hands help us learn. *Trends in Cognitive Sciences, 9*, 234-241.

Hegarty, M. (1992). Mental animation: Inferring motion from static displays of mechanical systems. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*, 1084-1102.

Hegarty, M., & Kozhevnikov, M. (1999). Types of visual-spatial representations and mathematical problem solving. *Journal of Educational Psychology, 91*, 684-689.

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

Kosslyn, S. M., Ball, T. M., & Reiser, B. J. (1978). Visual Images Preserve Metric Spatial Information - Evidence from Studies of Image Scanning. *Journal of Experimental Psychology-Human Perception and Performance, 4*, 47-60.

Kozhevnikov, M., Hegarty, M., & Mayer, R. E. (2002). Revising the visualizer-verbalizer dimension: Evidence for two types of visualizers. *Cognition and Instruction, 20*, 47-77.

Kozhevnikov, M., Kosslyn, S. M., & Shephard, J. (2005). Spatial versus object visualizers: a new characterization of visual cognitive style. *Memory & cognition, 33*, 710-726.

Kulvicki, J. V. (2013). *Images* (1 edition . ed.). New York: Routledge.

Larkin, J. H., & Simon, H. A. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science, 11*, 65-99.

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

Marr, D. C. (1982). *Vision: A computation investigation into the human representational system and processing of visual information*. San Francisco: Freeman.

Maywood, E. S., Reddy, A. B., Wong, G. K. Y., 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, 16*, 599-605.

Pinker, S. (1990). A theory of graph comprehension. In R. Feedle (Ed.), *Artificial Intelligence and the future of testing* (pp. 73-126). London: Psychology Press.

Reed, S., & Johnsen, J. (1975). Detection of parts in patterns and images. *Memory & cognition, 3*, 569-575.

Schwartz, D. L., & Black, J. B. (1996). Shuttling between depictive models and abstract rules: Induction and fallback. *Cognitive Science, 20*, 457-497.

Schwartz, D. L., & Black, T. (1999). Inferences through imagined actions: Knowing by simulated doing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 116-136.

Shah, P., & Carpenter, P. A. (1995). Conceptual limitations in comprehending line graphs. *Journal of Experimental Psychology-General, 124*, 43-61.

Shah, P., & Hoeffner, J. (2002). Review of graph comprehension research: Implications for instruction. *Educational Psychology Review, 14*, 47-69.

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science, 171*, 701-703.

Simcox, W. A. (1984). A method for pragmatic communication in graphic displays. *Human Factors, 26*, 483-487.

Stieff, M., Hegarty, M., & Dixon, B. (2010). Alternative strategies for spatial reasoning with diagrams. In A. Goel, M. Jamnik & N. H. Narayanan (Eds.), *Diagrammatic Representation and Inference* (Vol. 6170, pp. 115-127): Springer Berlin Heidelberg.

Takahashi, J. S., Hong, H.-K., Ko, C. H., & McDearmon, E. L. (2008). The genetics of mammalian circadian order and disorder: implications for physiology and disease. *Nat Rev Genet, 9*, 764-775.

Trickett, S., & Trafton, J. (2006). Toward a comprehensive model of graph comprehension: Making the case for spatial cognition. In D. Barker-Plummer, R. Cox & N. Swoboda (Eds.), *Diagrammatic representation and inference* (Vol. 4045, pp. 286-300): Springer Berlin / Heidelberg.

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

Ungerleider, L. G., & Mishkin, M. (1982). Two cortical visual systems. In D. J. Ingle, M. A. Goodale & R. J. W. Mansfield (Eds.), *Analysis of Visual Behavior* (pp. 549-586). Cambridge, MA: MIT Press.

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

Verstijnen, I. M., van Leeuwen, C., Goldschmidt, G., Hamel, R., & Hennessey, J. M. (1998). Creative discovery in imagery and perception: Combining is relatively easy, restructuring takes a sketch. *Acta Psychologica, 99*, 177-200.

Verstijnen, I. M., Van Leeuwen, C., Hamel, R., & Hennessey, J. M. (2000). What immagery can't do and why sketching might help. *Empirical Studies of the Arts, 18*, 167-182.

Welsh, D. K., Logothetis, D. E., Meister, M., & Reppert, S. M. (1995). Individual neurons dissociated from rat suprachiasmatic nucleus express independently phased circadian firing rhythms. *Neuron, 14*, 697-706.

Zacks, J. M., Levy, E., Tversky, B., & Schiano, D. J. (2002). Graphs in print. In M. Anderson, B. Meyer & P. Olivier (Eds.), *Diagrammatic representation and reasoning* (pp. 187-206). London: Springer.

Zacks, J. M., & Tversky, B. (1999). Bars and lines: a study of graphic communication. *Memory and Cognition, 27*, 1073-1079.

Zhang, J. (1997). The nature of external representations in problem solving. *Cognitive Science, 21*, 179-217.

Zhang, J., & Norman, D. A. (1994). Representations in distributed cognitive tasks. *Cognitive Science, 18*, 87-122.