CHAPTER 11

Thinking with Visualizations

One way to approach the design of an information system is to consider the *cost of knowledge*. Pirolli and Card (1995) drew an analogy with the way animals seek food to gain insights about how people seek information. Animals minimize energy expenditure to get the required gain in sustenance; humans minimize effort to get the necessary gain in information. Foraging for food has much in common with the seeking of information because, like edible plants in the wild, morsels of information are often grouped, but separated by long distances in an information wasteland. Pirolli and Card elaborated the idea to include information "scent"—like the scent of food, this is the information in the current environment that will assist us in finding more succulent information clusters.

The result of this approach is a kind of cognitive information economics. Activities are analyzed according to the value of what is gained and the cost incurred. There are two kinds of costs: resource costs and opportunity costs (Pirolli, 2003). *Resource costs* are the expenditures of time and cognitive effort incurred. *Opportunity costs* are the benefits that could be gained by engaging in other activities. For example, if we were not seeking information about information visualization, we might profitably be working on software design.

In some ways, an interactive visualization can be considered an *internal* interface between human and computer components in a problem-solving system. We are all becoming cognitive cyborgs in the sense that a person with a computer-aided design program, access to the Internet, and other software tools is capable of problem-solving strategies that would be impossible for that person acting unaided. A businessman plotting projections based on a spreadsheet business model can combine business knowledge with the computational power of the spreadsheet to plot scenarios rapidly, interpret trends visually, and make better decisions.

In this chapter, our concern is with the economics of cognition and the cognitive cost of knowledge. Human attention is a very limited resource. If it is taken up with irrelevant visual noise, or if the rate at which visual information is presented on the screen poorly matches the rate at which people can process visual patterns, then the system will not function well.

There are two fundamental ways in which visualizations support thinking: first, by supporting *visual queries* on information graphics, and second, by extending memory. For visual queries to be useful, the problem must first be cast in the form of a query pattern that, if seen, helps solve part of the problem. For example, finding a number of big red circles in a GIS display may indicate a problem with water pollution. Finding a long, red, fairly straight line on a map can show the best way to drive between two cities. Once the visual query is constructed, a visual search strategy, through eye movements and attention to relevant patterns, provides answers.

Memory extension comes from the way a display symbol, image, or pattern can rapidly evoke nonvisual information and cause it to be loaded from long-term memory into verbal-propositional processing centers.

This chapter presents the theory of how we think with visualizations. First, the memory and attention subsystems are described. Next, visual thinking is described as a set of embedded processes. Throughout, guidelines are provided for designing visual decision support systems.

Memory Systems

Memory provides the framework that underlies active cognition, whereas attention is the motor. As a first approximation, there are three types of memory: iconic, working, and long-term. There may also be a fourth, intermediate store that determines what from working memory finds its way into long-term memory. *Iconic memory* is a very brief image store, holding what is on the retina until it is replaced by something else or until several hundred milliseconds have passed (Sperling, 1960). *Long-term memory* is the information that we retain from everyday experience, perhaps for a lifetime. Consolidation of information into long-term memory only occurs, however, when active processing is done to integrate the new information with existing knowledge (Craik and Lockhart, 1972). *Visual working memory* holds the visual objects of immediate attention. These can be either external or mental images. In computer science terms, this is a register that holds information for the operations of visual cognition.

Visual Working Memory

The most critical cognitive resource for visual thinking is called *visual working memory*. Theorists disagree on details of exactly how visual working memory operates, but there is broad agreement on basic functionality and capacity—enough to provide a solid foundation for a theory of visual thinking. Closely related alternative concepts are the visuospatial sketchpad (Marr, 1982), visual short-term memory (Irwin, 1992), and visual attention (Rensink, 2002). Here is a list of some key properties of visual working memory:

- Visual working memory is separate from verbal working memory.
- Capacity is limited to a small number of simple visual objects and patterns, perhaps three to five simple objects.

- Positions of objects are stored in an egocentric map. Perhaps nine locations are stored, but only three to five are linked to specific objects.
- Attention controls what visual information is held and stored.
- The time to change attention is about 100 msec.
- The semantic meaning or gist of an object or scene (related more to verbal working memory) can be activated in about 100 msec.
- For items to be processed into long-term memory, deeper semantic coding is needed.

Working memory is not a single system; rather, it has a number of interlinked but separate components. There are separate systems for processing auditory and visual information, as well as subsystems for body movements and verbal output (Thomas et al., 1999). There may be additional stores for sequences of cognitive instructions and for motor control of the body. Kieras and Meyer (1997), for example, proposed an *amodal control memory*, containing the operations needed to accomplish current goals, and a general-purpose working memory, containing other miscellaneous information. A similar control structure is called the *central executive* in Baddeley and Hitch's model (1974), illustrated in Figure 11.1.

A detailed discussion of nonvisual working memory processes is beyond the scope of this book. Complete overview models of cognitive processes, containing both visual and nonvisual subsystems can be found in the Anderson ACT-R model (Anderson et al., 1997) and the executive process interactive control (EPIC) developed by Kieras and his coworkers (Kieras and Meyer, 1997). The EPIC architecture is illustrated in Figure 11.2. Summaries of the various working memory theories can be found in Miyake and Shah (1999).

That visual thinking results from the interplay of visual and nonvisual memory systems cannot be ignored. However, rather than getting bogged down in various theoretical debates about particular nonvisual processes, which are irrelevant to the perceptual issues, we will hereafter refer to nonvisual processes generically as *verbal-propositional* processing.

It is functionally quite easy to separate visual and verbal-propositional processing. Verbal-propositional subsystems are occupied when we speak, whereas visual subsystems are not. This allows for simple experiments to separate the two processes. Postma and De Haan (1996) provide a good example. They asked subjects to remember the locations of a set of easily recognizable objects—small pictures of cats, horses, cups, chairs, tables, etc.—laid out in two dimensions on a screen. Then the objects were placed in a line at the top of the display and the subjects were

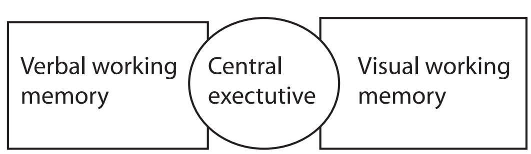

Figure 11.1 The multicomponent model of working memory of Baddeley and Hitch (1974).

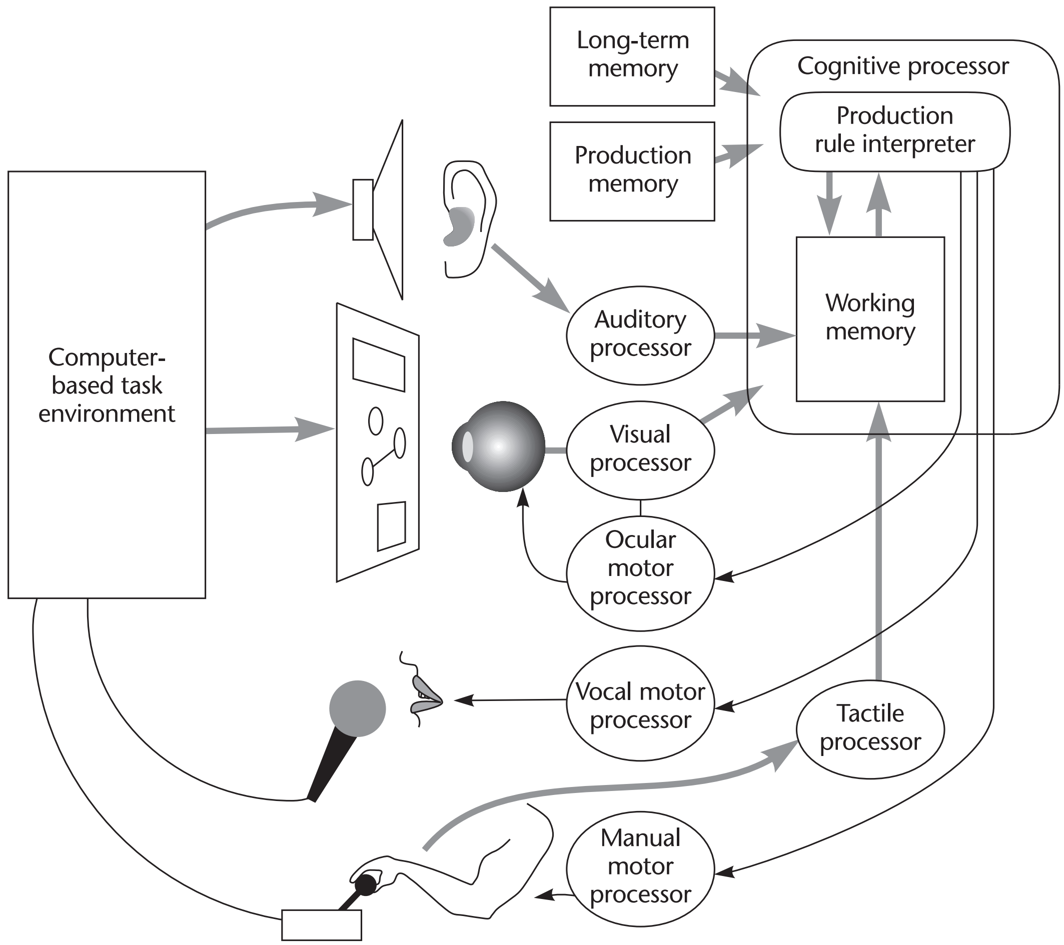

Figure 11.2 A unified extended cognitive model containing both human and machine processing systems. *Adapted from Kieras and Meyer, (1997).*

asked to reposition them—a task the subjects did quite well. In another condition, subjects were asked to repeat a nonsense syllable, such as *blah*, while in the learning phase; in this case they did much worse. However, saying *blah* did not disrupt memory for the *locations* themselves; it only disrupted memory for what was at the locations. This was demonstrated by having subjects place a set of disks at the positions of the original objects, which they could do with relative accuracy. In other words, when *blah* was said in the learning phase, subjects learned a set of locations but not the objects at those locations. This technique is called *articulatory suppression* (Postma and DeHaan, 1996; Postma et al., 1998).

Presumably, the reason why saying *blah* disrupted memory for the objects is that this information was translated into a verbal-propositional coding when the objects were attended.

Visual Working Memory Capacity

Position is not the only information stored in visual working memory; some abstract shape, color, and texture information is also retained. Visual working memory can be roughly defined as the visual information retained from one fixation to the next. This appears to be limited to about three to five simple objects (Irwin, 1992; Xu, 2002; Luck and Vogel 1997; Melcher, 2001). The exact number depends on the task and the kind of pattern. Figure 11.3(a) illustrates the kinds of patterns used in a series of experiments by Vogel et al. (2001). In these experiments, one set of objects was shown for a fraction of a second (e.g., 400 msec), followed by a blank of more than 0.5 sec. After the blank, the same pattern was shown, but with one attribute of an object altered—for example, its color or shape. The results from this and a large number of similar studies have shown that about three objects can be retained without error, but these objects can have color, shape, and texture. If the same amount of color, shape, and texture information is distributed across more objects, memory declines for each of the attributes.

Only quite simple shapes can be stored in this way. For example, each of the mushroom shapes shown in Figure 11.3(b) uses up two visual memory slots (Xu, 2002). Subjects do no better if the stem and the cap are combined than if they are separated. Intriguingly, Vogel et al. (2001) found that if colors were combined with concentric squares, as shown in Figure 11.3(c), then six colors could be held in visual working memory, but if they were put in side-by-side squares, only three colors could be retained. Melcher (2001) found that more information could be retained if longer viewing was permitted: up to five objects after a four-second presentation.

What are the implications for data glyph design? (A *glyph*, as discussed in Chapter 5, is a visual object that displays one or more data variables.) If it is important that a data glyph be held in visual working memory, then it is important that its shape allows it to be encoded according to visual working memory capacity. For example, Figure 11.4 shows two ways of representing the same data. One consists of an integrated glyph containing a colored arrow showing orientation, by arrow direction; temperature, by arrow color; and pressure, by arrow width. A second representation distributes the three quantities among three separate visual objects: orientation by

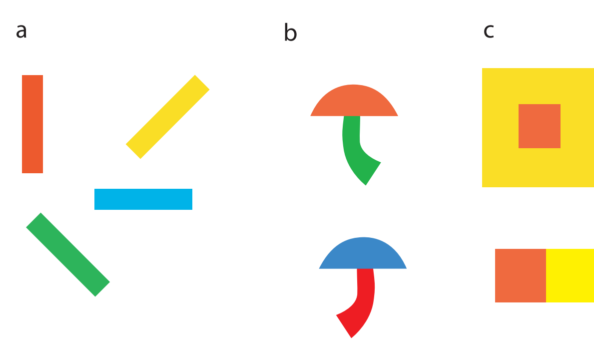

Figure 11.3 Patterns used in studies of the capacity of visual working memory. *(a) From Vogel et al. (2001). (b) From Xu (2002). (c) From Vogel et al. (2001).*

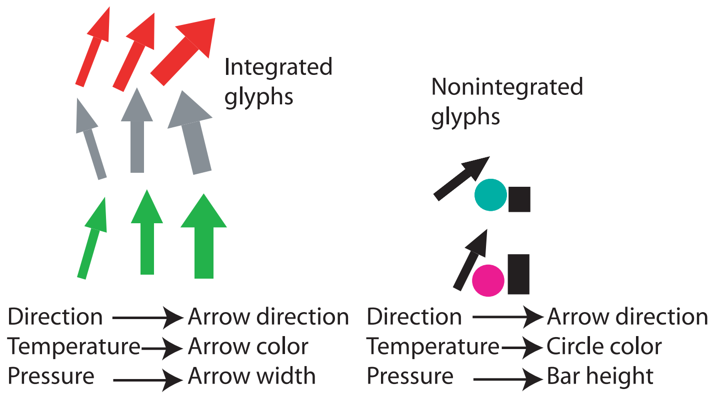

Figure 11.4 If multiple data attributes are integrated into a single glyph, more information can be held in visual working memory.

an arrow, temperature by the color of a circle, and air pressure by the height of a rectangle. The theory of visual working memory and the results of Vogel et al. (2001) suggest that three of the former glyphs could be held in visual working memory, but only one of the latter.

Object Files, Coherence Fields, and Gist

What exactly is held in working memory? Kahneman et al. (1992) coined the term *object file* to describe the temporary grouping of a collection of visual features together with other links to verbal-propositional information. They hypothesized that an object file would consist of a neural activation pattern having the equivalent of pointers reaching into the part of the brain where visual features are processed, as well as pointers to verbal working memory structures and to stored motor memories concerned with the appropriate body movements to make in response.

What we perceive is mostly determined by the task at hand, whether it is finding a path over rocks or finding the lettuce in a grocery store. Perception is tuned by the task requirements to give us what is most likely to be useful. In the first example we see the rocks immediately in front of us. In the second we see green things on the shelves. We can think of perception as occurring through a sequence of active *visual queries* operating through a focusing of attention to give us what we need. The neural mechanism underlying the query may be a rapid tuning of the pattern perception networks to respond best to patterns of interest (Dickinson et al. 1997). Rensink (2002, 2000) coined the term *nexus* to describe this instantaneous grouping of information by attentional processing.

Another term sometimes used to describe a kind of summary of the properties of an object or a scene is *gist*. Gist is used mainly to refer to the properties that are pulled from long-term memory as the image is recognized. Visual images can activate verbal-propositional information in as little as 100 msec (Potter, 1976). Gist consists of both visual information about the typical structure of an object and links to relevant verbal-propositional information. We may also store the gist of a whole environment, so that when we see a familiar scene, the interior of a car, for example, a whole visual framework of the typical locations of things will be activated. We can think of an object file as the temporary structure in working memory, whereas gist is a longer-term counterpart.

Change Blindness

One of the consequences of the very small amount of information held in visual working memory is a phenomenon known as *change blindness* (Rensink, 2000). Because we remember so little, it is possible to make large changes in a display between one view and the next and people generally will not notice, unless the change is to something they have recently attended. If a change is made while the display is being fixated, the inevitable blink will draw attention to it. But if changes are made mid–eye movement, midblink, or after a short blanking of the screen (Rensink, 2002), the change generally will not be seen. Iconic memory information in retinal coordinates decays within about 200 msec (Phillips, 1974). By the time 400 msec have elapsed, what little remains is in visual working memory.

An extraordinary example of change blindness is a failure to detect a change from one person to another in midconversation. Simons and Levin (1998) carried out a study in which an unsuspecting person was approached by a stranger holding a map and asking for directions. The conversation that ensued was interrupted by two workers carrying a door and during this interval another actor, wearing different clothes, was substituted to carry on the conversation. Remarkably, most people did not notice the substitution.

To many people, the extreme limitation on the capacity of visual working memory seems quite incredible. How can we experience a rich and detailed world, given such a shallow internal representation? The answer to this dilemma is that the world "is its own memory" (O'Regan, 1992). We perceive the world to be rich and detailed, not because we have an internal detailed model, but simply because whenever we wish to see detail we can get it, either by focusing attention on some aspect of the visual image at the current fixation or by moving our eyes to see the detail in some other part of the visual field. We are unaware of the jerky eye movements by which we explore the world; we are only aware of the complexity of the environment detail being brought into working memory on a need-to-know, just-in-time fashion (O'Regan, 1992; Rensink, 2002; Rensink et al., 1997). This is in agreement with the idea of visual queries being basic to perception.

Spatial Information

For objects acquired in one fixation to be reidentified in the next requires some kind of buffer that holds locations in egocentric coordinates as opposed to retina-centric coordinates (Hochberg, 1968). This also allows for the synthesis of information obtained from successive fixations. Figure 11.5 illustrates the concept. Neurophysiological evidence from animal studies suggests that the lateral interparietal area near the top of the brain (Colby, 1998) appears to play a crucial role in linking eye-centered coordinate maps in the brain with egocentric coordinate maps.

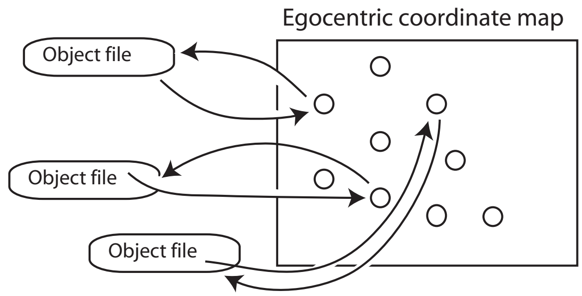

Figure 11.5 A spatial map of objects that have recently been held by attention is a necessary part of visual working memory.

Egocentric-spatial location memory also holds remarkably little information, although probably a bit more than the three objects that Vogel et al. (2001) suggest. It may be possible to remember some information about approximately nine locations (Postma et al., 1998). Three of these may contain links to object files, whereas the remaining ones specify only that there is something at a particular region in space, but very little more. Some evidence suggests that fixation of a particular object may be essential for that object and its location to be held from one fixation to the next (Hollingworth and Henderson, 2002).

Some visual information is retained over several seconds and several fixations. Potter (2002) provided evidence for this. Subjects viewed a rapid serial presentation of 10 pictures at the rate of six per second and afterwards were able to identify whether a particular picture was in the set about 60% of the time. This suggests that some residual gist is retained over many visual changes in scene. A recent and very intriguing study by Melcher (2001) suggests that we can build up information about several scenes that are interspersed. When the background of a scene was shown, subjects could recall some of the original objects, even though several other scenes had intervened. This implies that a distinctive screen design could help with visual working memory when we switch between different views of a data space. We may be able to cognitively swap in and swap out different data "scenes," albeit each with a low level of detail.

An interesting question is how many moving targets can be held from one fixation to the next. The answer seems to be about four or five. Pylyshyn and Storm (1988) carried out experiments in which visual objects moved around on a display in a pseudo-random fashion. A subset of the objects was visually marked by changing color, but then the marking was turned off. If there were five or fewer marked objects, subjects could continue to keep track of them, even though they were now all black. Pylyshyn coined the term *FINST*, for *fingers of instantiation*, to describe the set of pointers in a cognitive spatial map that would be necessary to support this task. The number of individual objects that can be tracked is somewhat larger than the three found by Vogel et al. (2001), although it is possible that the moving objects may be grouped perceptually into fewer chunks (Yantis, 1992).