Attention

The studies showing that we can hold three or four objects in visual working memory required intense concentration on the part of the participants. Most of the time, when we interact with displays or just go about our business in the everyday world, we will not be attending that closely. In a remarkable series of studies, Mack and Rock (1998) tricked subjects into not paying attention to the subject of the experiment, although they wanted to make sure that subjects were at least looking in the right direction. They told subjects to attend to a cross pattern for changes in the length of one of the arms; perfect scores on this task indicated they had to be attending. Then the researchers presented a pattern that the subject *had not been asked to look for*. They found that even though the unexpected pattern was close to, or even on, the point of fixation, most of the time it was not seen. The problem with this kind of study is that the ruse can only be used once. As soon as you ask subjects if they saw the unexpected pattern, they will start looking for unexpected patterns. Mack and Rock therefore used each subject for only one trial; they used literally hundreds of subjects in a series of studies.

Mack and Rock called the phenomenon *inattentional blindness*. It should not be considered as a peculiar effect only found in the laboratory. Instead this kind of result probably reflects everyday reality much more accurately than the typical psychological experiment in which subjects are paid to closely attend. Most of the time we simply do not register what is going on in our environment unless we are looking for it. The conclusion must be that attention is central to all perception.

Although we are blind to many changes in our environment, some visual events are more likely to cause us to change attention than others are. Mack and Rock found that although subjects were blind to small patterns that appeared and disappeared, they still noticed larger visual events, such as patterns larger than one degree of visual angle appearing near the point of fixation.

Jonides (1981) studied ways of moving a subject's attention from one part of a display to another. He looked at two different ways, which are sometimes called pull cues and push cues. In a *pull cue*, a new object appearing in the scene pulls attention toward it. In a *push cue*, a symbol in the display, such as an arrow, tells someone where a new pattern is to appear. It appears to take only about 100 msec to shift attention based on a pull cue but can take between 200 and 400 msec to shift attention based on a push cue.

Visual attention is not strictly tied to eye movements. Although attending to some particular part of a display often does involve an eye movement, there are also attention processes operating within each fixation. The studies of Triesman and Gormican (1988) and others (discussed in Chapter 5) showed that we process simple visual objects serially at a rate of about one every 40–50 msec. Because each fixation typically will last for 100–300 msec, this means that our visual systems process two to six objects within each fixation, before we move our eyes to attend visually to some other region.

Attention is also not limited to specific locations of a screen. We can, for example, choose to attend to a particular pattern that is a component of another pattern, even though the

patterns overlap spatially (Rock and Gutman, 1981). Thus, we can choose to attend to the curved pattern or to the rectangular shape in Figure 11.6. We can also choose to attend to a particular attribute if it is preattentively distinct (Treisman 1985). For example, on a field of black text with parts highlighted in red, we can choose to attend only to the red items. Having whole groups of objects that move is especially useful in helping us to attend selectively (Bartram and Ware, 2002). We can attend to the moving group or the static group, with relatively little interference between them.

The selectivity of attention is by no means perfect. Even though we may wish to focus on one aspect of a display, other information is also processed, apparently to quite a high level. The well known Stroop effect illustrates this (Stroop, 1935). In a set of words printed in different colors, as illustrated in Figure 11.7, if the words themselves are color names that do not match the ink colors, subjects name the colors more slowly than if the colors match the words. This means that the words are processed automatically; we cannot entirely ignore them even when we want to. More generally, it is an indication that all highly learned symbols will automatically invoke verbal-propositional information that has become associated with them.

The Role of Motion in Attracting Attention

As we conduct more of our work in front of computer screens, there is an increasing need for signals that can attract a user's attention. Often someone is busy with a primary task, perhaps filling out forms or composing email, while at the same time events may occur on other parts of the display, requiring attention. These *user interrupts* can alert us to an incoming message from

Figure 11.6 We can attend to either the curved orange shape or the black rectangle, even though they overlap in space.

RED GREEN YELLOW BLUE BLACK GREEN PURPLE BLUE BLACK
ORANGE GREEN RED GREEN YELLOW BLUE BLACK GREEN PURPLE
BLUE BLACK ORANGE BLACK GREEN RED

GREEN RED BLUE YELLOW PURPLE RED BLACK BLUE BLACK GREEN
ORANGE BLUE RED PURPLE YELLOW RED BLACK YELLOW GREEN
ORANGE BLACK GREEN RED GREEN

Figure 11.7 As quickly as you can, try to name the colors of the words in the set at the top. Then try to name the colors in the set below. Even though both tasks involve ignoring the words themselves, people are slowed down by the mismatch. This is called the Stroop effect.

a valued customer or a signal from a computer agent that has been out searching the Internet for information on the latest flu virus. There are four basic visual requirements for a user interrupt:

- A signal should be easily perceived, even if it is outside of the area of immediate focal attention.
- If the user wishes to ignore the signal and attend to another task, the signal should continue to act as a reminder.
- The signal should not be so irritating that it makes the computer unpleasant to use.
- It should be possible to endow the signal with various levels of urgency.

Essentially, the problem is how to attract the user's attention to information outside the central parafoveal region of vision (approximately the central six degrees). For a number of reasons, the options are limited. We have a low ability to detect small targets in the periphery of the visual field. Peripheral vision is color blind, which rules out color signals (Wyszecki and Stiles, 1982). Saccadic suppression during eye movements means that some transitory event occurring in the periphery will generally be missed if it occurs during an eye movement (Burr and Ross, 1982). Taken together, these facts suggest that a single, abrupt change in the appearance of an icon is unlikely to be an effective signal.

The set of requirements suggests two possible solutions. One is to use auditory cues. In certain cases, these are a good solution, but they are outside the scope of this book. Another solution is to use blinking or moving icons. In a study involving shipboard alarm systems, Goldstein and Lamb (1967) showed that subjects were capable of distinguishing five flash patterns with approximately 98% reliability and that they responded with an average delay of approximately 2.0 seconds. But anecdotal evidence indicates that a possible disadvantage of flashing lights or blinking cursors is that users find them irritating. Unfortunately, many Web page designers generate a kind of animated chart junk: small, blinking animations with no functional purpose are often used to jazz up a page. Moving icons may be a better solution. Moving targets are detected more easily in the periphery than static targets (Peterson and Dugas, 1972). In a series of dual task experiments, Bartram et al. (2003) had subjects carry out a primary task, either text editing or playing Tetris or Solitaire, while simultaneously monitoring for a change in an icon at the side of the display in the periphery of the visual field. The results showed that having an icon move was far more effective in attracting a user's attention than having it change color or shape. The advantage increased as the signal was farther from the focus of attention in the primary task.

Another advantage of moving or blinking signals is that they can persistently attract attention, unlike a change in an icon, such as the raising of a mailbox flag, which fades rapidly from attention. Also, although rapid motions are annoying, slower motions need not be and they can still support a low-level of awareness (Ware et al., 1992).

Interestingly, more recent work has suggested that it may not be motion per se that attracts attention, but the appearance of a new object in the visual field (Hillstrom and Yantis, 1994;

Enns et al., 2001). This seems right; after all, we are not constantly distracted in an environment of swaying trees or people moving about on a dance floor. It also makes ecological sense; when early man was outside a cave, intently chipping a lump of flint into a hand axe, or when early woman was gathering roots out on the grassland, awareness of emerging objects in the periphery of vision would have had clear survival value. Such a movement might have signaled an imminent attack. Of course, the evolutionary advantage goes back much further than this. Monitoring the periphery of vision for moving predators or prey would provide a survival advantage for most animals. Thus, the most effective reminder might be an object that moves into view, disappears, and then reappears every so often. In a study that measured the eye movements made while viewing multimedia presentations, Faraday and Sutcliffe (1997) found that the onset of motion of an object generally produced a shift of attention to that object.

Rensink's Model

Rensink (2002) has recently developed a model that ties together many of the components we have been discussing. Figure 11.8 illustrates. At the lowest level are the elementary visual features that are processed in parallel and automatically. These correspond to elements of color, edges, motion, and stereoscopic depth. From these elements, prior to focused attention, low-level precursors of objects, called *proto-objects*, exist in a continual state of flux. At the top level, the mechanism of attention forms different visual objects from the proto-object flux. Note that Rensink's proto-objects are located at the top of his "low-level vision system." He is not very specific on the nature of proto-objects, but it seems reasonable to suppose that they have char-

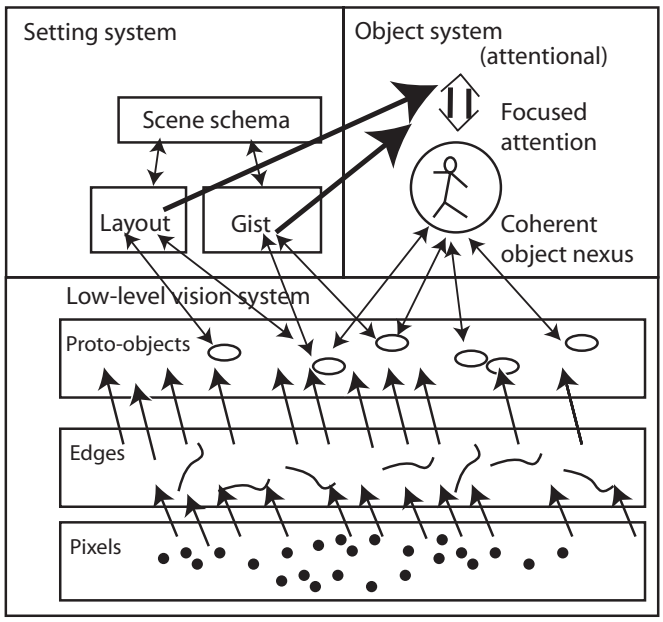

Figure 11.8 A model of visual attention. *Adapted from Rensink (2002).*

acteristics similar to the mid-level pattern perception processes in the three-stage model laid out in this book.

Rensink uses the metaphor of a hand to represent attention, with the fingers reaching down into the proto-object field to instantiate a short-lived object. After the grasp of attention is released, the object loses its coherence and the components fall back into the constituent proto-objects. There is little or no residue from this attentional process. Other components of the model are a layout map containing location information and the rapid activation of object gist.

The central role of attention in Rensink's model suggests a way that visual queries can be used to modify the grasp of attention and pull out the particular patterns we need to support problem solving. For example, we might need to know how one module connects to another in a software system. To obtain this information, a visual query is constructed to find out if lines connect certain boxes in the diagram. This query is executed by focusing visual attention on those graphical features.

The notion of proto-objects in a continuous state of flux suggests, also, how visual displays can provide a basis for creative thinking, because they allow multiple visual interpretations drawn from the same visualization. Another way to think about this is that different patterns in the display become cognitively highlighted, as we consider different aspects of a problem.

Eye Movements

We constantly make eye movements to seek information. Moving our eyes causes different parts of the visual environment to be imaged on the high-resolution fovea, where we can see detail. These movements are frequent. For example, as you read this page, your eye is making between two and five jerky movements, called *saccades*, per second.

Here are the basic statistics describing three important types of eye movement:

1. Saccadic movements: In a visual search task, the eye moves rapidly from fixation to fixation. The dwell period is generally between 200 and 600 msec, and the saccade takes between 20 and 100 msec. The peak velocity of a saccade can be as much as 900 deg/sec (Hallett, 1986; Barfield et al., 1995).

2. Smooth-pursuit movements: When an object is moving smoothly in the visual field, the eye has the ability to lock onto it and track it. This is called a *smooth-pursuit* eye movement. This ability also enables us to make head and body movements while maintaining fixation on an object of interest.

3. Convergent movements (also called *vergence* movements): When an object moves toward us, our eyes converge. When it moves away, they diverge. Convergent movements can be either saccadic or smooth.

Saccadic eye movements are said to be *ballistic*. This means that once the brain decides to switch attention and make an eye movement, the muscle signals for accelerating and decelerating the

eye are first programmed, then the program is run to make the eye movement. The movement cannot be adjusted in mid-saccade. During the course of a saccadic eye movement, we are less sensitive to visual input than we normally are. This is called *saccadic suppression* (Riggs et al., 1974). The implication is that certain kinds of events can easily be missed if they occur while we happen to be moving our eyes. This is important when we consider the problem of alerting a computer operator to an event.

Another implication of saccadic suppression is that it is reasonable to think of information coming into the visual system as a series of discrete snapshots. The brain is often processing rapid sequences of discrete images. This capacity is being increasingly exploited in television advertising, in which several cuts per second of video have become commonplace.

Accommodation

When the eye moves to a new target at a different distance from the observer, it must refocus, or *accommodate*, so that the target is clearly imaged on the retina. An accommodation response typically takes about 200 msec. The mechanisms controlling accommodation and convergent eye movements are neurologically coupled, and this can cause problems with virtual-reality displays. This problem is discussed in Chapter 8.

Eye Movements, Search, and Monitoring

How does the brain plan a sequence of eye movements to interpret a visual scene? A simple heuristic strategy appears to be employed according to the theory of Wolfe and Gancarz (1996). First, the feature map of the entire visual field is processed in parallel (see Chapter 5) to generate a map weighted according to the current task. For example, if we are scanning a crowd to look for people we know, the feature set will be highly correlated with human faces. Next, eye movements are executed in sequence, visiting the strongest possible target first and proceeding to the weakest. Finally once each area has been processed, it is cognitively flagged as visited. This has the effect of inhibiting that area of the weighted feature map.

A searchlight is a useful metaphor for describing the interrelationships among visual attention, eye movements, and the useful field of view. In this metaphor, visual attention is like a searchlight used to seek information. We point our eyes at the things we want to attend to. The diameter of the searchlight beam, measured as a visual angle, describes the useful field of view (UFOV). The central two degrees of visual angle is the most useful, but it can be broader, depending on such factors as stress level and task. The direction of the searchlight beam is controlled by eye movements. Figure 11.9 illustrates the searchlight model of attention.

Supervisory Control

The searchlight model of attention has been developed mainly in the context of supervisory control systems to account for the way people scan instrument panels. *Supervisory control* is a term used for complex, semiautonomous systems that are only indirectly controlled by human operators.

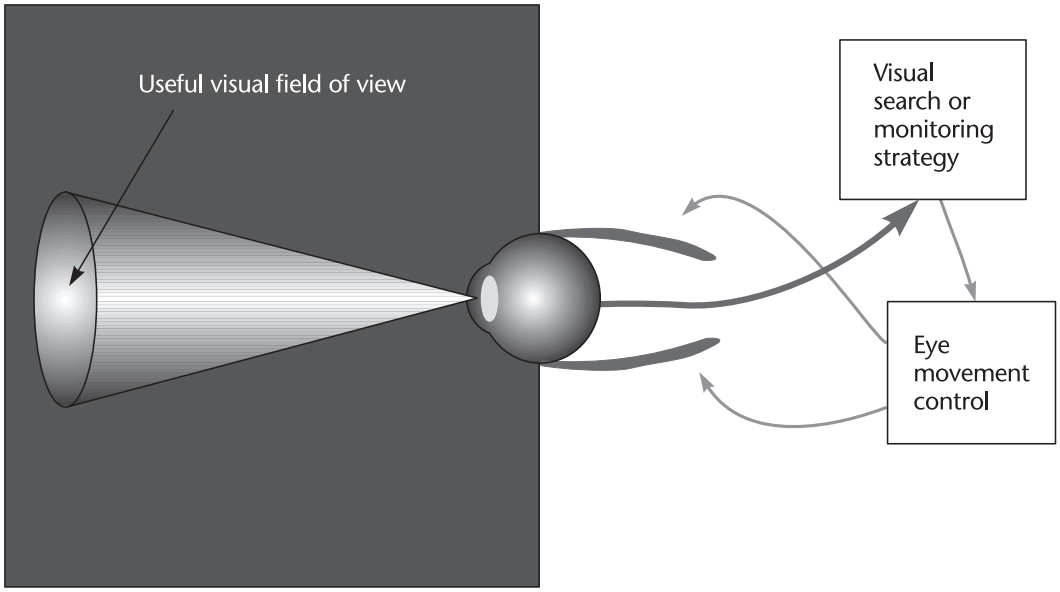

Figure 11.9 The searchlight model of attention. A visual search strategy is used to determine eye movements that bring different parts of the visual field into the useful visual field of view.

Examples are sophisticated aircraft and power stations. In these systems, the human operator has both a monitoring and a controlling role. Because the consequences of making an error during an emergency can be truly catastrophic, a good interface design is critical. Two Airbus passenger jets have crashed for reasons that are attributed to mistaken assumptions about the behavior of supervisory control systems (Casey, 1993). There are also stories of pilots in fighter aircraft turning warning lights off because they are unable to concentrate in a tense situation.

A number of aspects of visual attention are important when considering supervisory control. One is creating effective ways for a computer to gain the attention of a human—a user-interrupt signal. Sometimes, a computer must alert the operator with a warning of some kind, or it must draw the operator's attention to a routine change of status. In other cases, it is important for an operator to become aware of *patterns* of events. For example, on a power grid, certain combinations of component failures can indicate a wider problem. Because display panels for power grids can be very large, this may require the synthesis of widely separated visual information.

In many ways, the ordinary human–computer interface is becoming more like a supervisory control system. The user is typically involved in some foreground task, such as preparing a report, but at the same time monitoring activities occurring in other parts of the screen.

Visual Monitoring Strategies

In many supervisory control systems, operators must monitor a set of instruments in a semi-repetitive pattern. Models developed to account for operators' visual scanning strategies generally have the following elements (Wickens, 1992):

Channels: These are the different ways in which the operator can receive information. Channels can be display windows, dials on an instrument panel, or nonvisual outputs, such as loudspeakers (used for auditory warnings).

Events: These are the signals occurring on channels that provide useful information.

Expected cost: This is the cost of *missing* an event.

System operators base their monitoring of different channels on a mental model of system event probabilities and the expected costs of these (Moray and Rotenberg, 1989; Wickens, 1992). Charbonnell et al. (1968) and Sheridan (1972) proposed that monitoring behavior is controlled by two factors: the growth of uncertainty in the state of a channel (between samples) and the cost of sampling a channel. *Sampling* a channel involves fixating part of a display and extracting the useful information. The cost of sampling is inversely proportional to the ease with which the display can be interpreted. This model has been successfully applied by Charbonnell et al. (1968) to the fixation patterns of pilots making an instrument landing. A number of other factors may influence visual scanning patterns:

- Operators may minimize eye movements. The cost of sampling is reduced if the points to be sampled are spatially close. Russo and Rosen (1975) found that subjects tended to make comparisons most often between spatially adjacent data. If two indicators are within the same effective field of view, this tendency will be especially advantageous.
- There can be oversampling of channels on which infrequent information appears (Moray, 1981). This can be accounted for by short-term memory limitations. Human working memory has very limited capacity, and it requires significant cognitive effort to keep a particular task in mind. People can reliably monitor an information channel every minute or so, but they are much less reliable when asked to monitor an event every 20 minutes. One design solution is to build in visual or auditory reminders at appropriate intervals.
- Sometimes operators exhibit dysfunctional behaviors in high-stress situations. Moray and Rotenberg (1989) suggested that under crisis conditions, operators cease monitoring some channels altogether. In an examination of control-room emergency behavior, he found that under certain circumstances, an operator's fixation became locked on a feedback indicator, waiting for a system response at the expense of taking other, more pressing actions.
- Sometimes, operators exhibit systematic scan patterns, such as the left-to-right, top-to-bottom one found in reading, even if these have no functional relevance to the task (Megaw and Richardson, 1979).

Long-Term Memory

We now turn away from strictly visual processing to consider the structure of information in verbal-propositional memory. We will need this background information to understand how visualizations can function as memory aids by rapidly activating structured nonvisual information.

Long-term memory contains the information that we build up over a lifetime. We tend to associate long-term memory with events we can consciously recall—this is called *episodic memory* (Tulving, 1983). However, long-term memory also includes motor skills, such as the finger movements involved in typing and the perceptual skills, integral to our visual systems, that enable us to rapidly identify words and thousands of visual objects. Nonvisual information that is not closely associated with concepts currently in verbal working memory can take minutes, hours, or days to retrieve from long-term memory.

There is a common myth that we remember everything we experience but we lose the indexing information; in fact, we remember only what gets encoded in the first 24 hours or so after an event occurs. The best estimates suggest that we do not actually store very much information in long-term memory. Using a reasonable set of assumptions, Landauer (1986) estimated that only about 10^9 bits of information are stored over a 35-year period. This is what can currently be found in the solid-state main memory of a personal computer. The power of human long-term memory is not in its capacity but in its remarkable flexibility. The same information can be combined in many different ways and through many different kinds of cognitive operations.

Human long-term memory can be usefully characterized as a network of linked concepts (Collins and Loftus, 1975; Yufic and Sheridan, 1996). Our intuition supports this model. If we think of a particular concept—for example, data visualization—we can easily bring to mind a set of related concepts: computer graphics, perception, data analysis, potential applications. Each of these concepts is linked to many others. Figure 11.10 shows some of the concepts relating to information visualization.

The network model makes it clear why some ideas are harder to recall than others. Concepts and ideas that are distantly related naturally take longer to find; it can be difficult to trace a path to them and easy to take wrong turns in traversing the concept net, because no map exists. For this reason, it can take minutes, hours, or even days to retrieve some ideas. A study by Williams and Hollan (1981) investigated how people recalled names of classmates from their high school graduating class, seven years later. They continued to recall names for at least 10 hours, although the number of falsely remembered names also increased over time. The forgetting of information from long-term memory is thought to be more of a loss of access than an erasure of the memory trace (Tulving and Madigan, 1970). Memory connections can easily become corrupted or misdirected; as a result, people often misremember events with a strong feeling of subjective certainty (Loftus and Hoffman, 1989).

Chunks of information are continuously being prioritized, and to some extent reorganized, based on the current cognitive requirements (Anderson and Milson, 1989). It is much easier to recall something that we have recently had in working memory. Seeing an image from the past will prime subsequent recognition so that we can identify it more rapidly (Bichot and Schall, 1999).

Long-term memory and working memory appear to be overlapping, distributed, and specialized. Long-term visual memory involves parts of the visual cortex, and long-term verbal memory involves parts of the temporal cortex specialized for speech. More abstract and linking concepts may be represented in areas such as the prefrontal cortex. Working memory is better

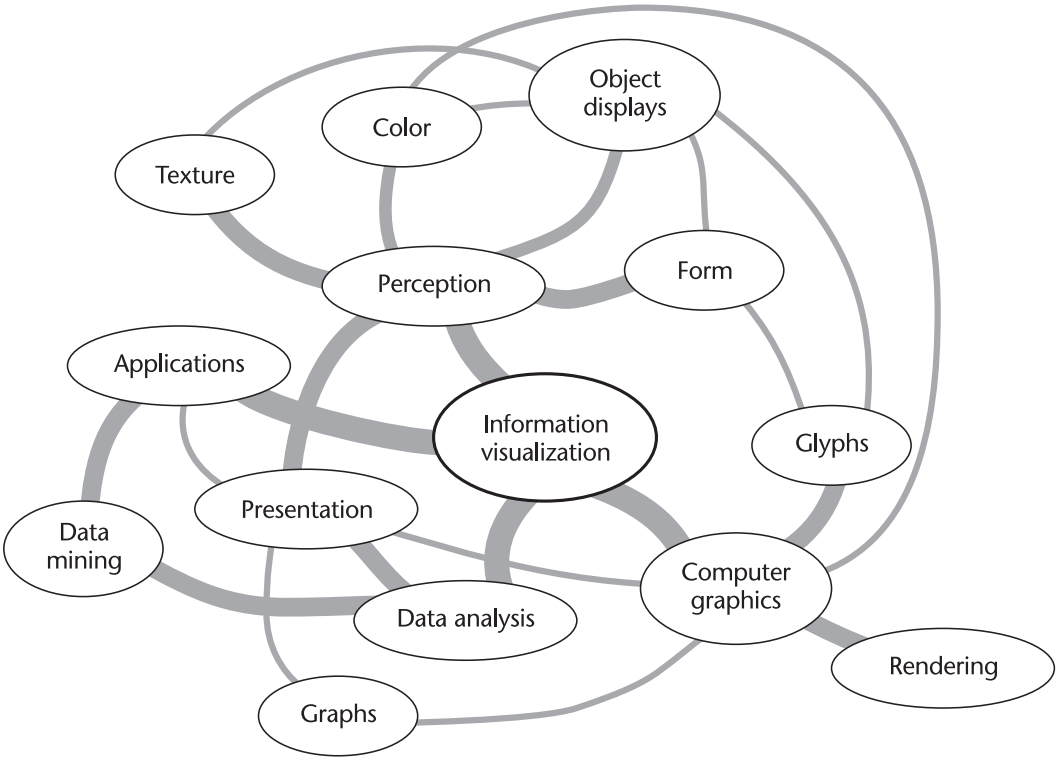

Figure 11.10 A concept map showing a set of linked concepts surrounding the idea of information visualization.

thought of as existing within the context of long-term memory than as a distinct processing module. As visual information is processed through the visual system, it activates the long-term memory of visual objects that have previously been processed by the same system. This explains why visual recognition is much faster and more efficient than visual recall. In recognition, a visual memory trace is being reawakened, so that we know that we have seen a particular pattern. In recall, it is necessary for us actually to describe some pattern, by drawing or in words, but we may not have access to the memory trace. In any case, the memory trace will not generally contain sufficient information for reconstructing an object, which would be required for recognition but not for recall. The memory trace also explains priming effects: If a particular neural circuit has recently been activated, it becomes primed for reactivation.

Chunks and Concepts

Human memory is much more than a simple repository like a telephone book; information is highly structured in overlapping and interconnected ways. The term *chunk* and the term *concept* are both used in cognitive psychology to denote important units of stored information. The two terms are used interchangeably here. The process of grouping simple concepts into more complex ones is called *chunking*. A chunk can be almost anything: a mental representation of

an object, a plan, a group of objects, or a method for achieving some goal. The process of becoming an expert in a particular domain is largely one of creating effective high-level concepts or chunks.

It is generally thought that concepts are formed by a kind of hypothesis-testing process (Levine, 1975). According to this view, multiple tentative hypotheses about the structure of the world are constantly being evaluated based on sensory evidence and evidence from internal long-term memory. In many cases, the initial hypotheses start with some existing concept: a mental model or metaphor. New concepts are distinguished from the prototype by means of transformations (Posner and Keele, 1968). For example, the concept of a zebra can be formed from the concept of a horse by adding a new node to a concept net containing a reference to a horse and distinguishing information, such as the addition of stripes.

What about purely visual long-term memory? It does not appear to contain the same kind of network of abstract concepts that characterizes verbal long-term memory. However, there may be some rather specialized structures in visual scene memory. Evidence for this comes from studies showing that we identify objects more rapidly in the right context, such as bread in a kitchen (Palmer, 1975). The power of images is that they rapidly evoke verbal-propositional memory traces; we see a cat and a whole host of concepts associated with cats becomes activated. Images provide rapid evocation of the semantic network, rather then forming their own net (Intraub and Hoffman, 1992). To identify all of the objects in our visual environment requires a great store of visual appearance information. Biederman (1987) estimated that we may have about 30,000 categories of visual information. The way visual objects are cognitively constructed is discussed more extensively in Chapter 8.

Visual imagery is the basis for a well-known mnemonic technique called the *method of loci* (Yates, 1966). This was known to Greek and Roman orators and can be found in many self-help books on how to improve your memory. To use the method of loci, you must identify a path that you know well, such as the walk from your house to the corner store. To remember a series of words—for example, mouse, bowl, fork, box, scissors—place each object at specific locations along the path in your mind's eye. You might put one at the end of your driveway next to the mailbox, the next by a particular lamppost, and so on. Now, to recall the sequence, you simply take an imaginary walk—magically, the objects are where you have placed them. The fact that this rather strange technique actually works suggests that there is something special about associating concepts to be remembered with images in particular locations that helps us remember information.

The Data Mountain was an experimental computer interface designed to take advantage of the apparent mnemonic value of spatial layout (Robertson et al., 1998). The Data Mountain allowed users to lay out thumbnails of Web pages on the slope of an inclined plane, as illustrated in Figure 8.4. A study by Czerwinski et al. (1999) found that even six months later, subjects who had previously set up information in this way could find particular items as rapidly as they could shortly after the initial layout. It should be noted, however, that before retesting subjects were given a practice session that allowed them to relearn at least some of the layout; it is possible to scan a lot of information in the two minutes or so that they were given.

Using a setup very similar to the Data Mountain, Cockburn and Mackenzie (2001) showed that removing the perspective distortion, as shown in Figure 8.4, had no detrimental impact on performance. Thus, although spatial layout may aid memory, it does not, apparently, have to be a 3D layout. On balance, there does appear to be support for the mnenomic value of spatial layout, because a lot of items (i.e., 100 items) were used in the memory test and the review period was brief, but there is little evidence that the space must be three-dimensional.

One important aspect of the Data Mountain study was that subjects were required to organize the material into categories. This presumably caused a deeper level of cognitive processing. Depth of processing is considered a primary factor in the formation of long-term memories (Craik and Lockhart, 1972). To learn new information, it is not sufficient to be exposed to it over an over again, the information must be integrated cognitively with existing information. Tying verbal and visual concepts together may be especially effective. Indeed, this is a central premise in the use of multimedia in education.

Problem Solving with Visualizations

We are now in a position to outline a theory of how thinking can be augmented by visual queries on visualizations of data. Figure 11.11 provides an overview of the various components. This

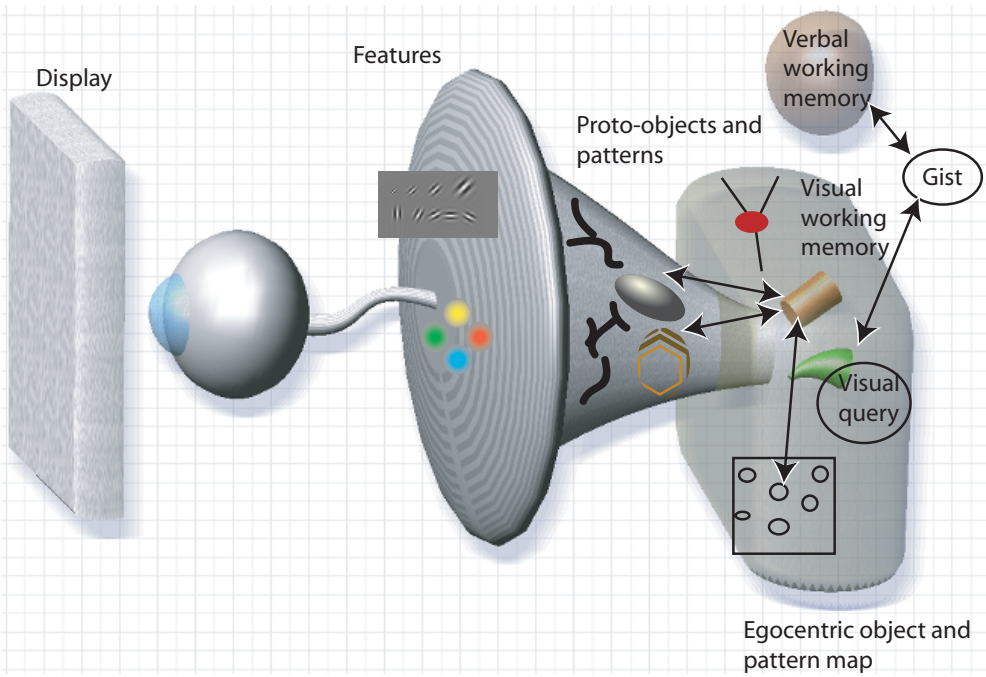

Figure 11.11 The cognitive components involved in visual thinking.

borrows a great deal from Rensink (2000, 2002) and earlier theorists, such as Baddeley and Hitch (1974), Kahneman et al. (1992) and Triesman (1985). It is based on the three-stage model developed throughout this book, but with the third stage now elaborated as an active process. At the lowest stage is the massively parallel processing of the visual scene into the elements of form, opponent colors, and the elements of texture and motion. In the middle stage is pattern formation, providing the basis for object and pattern perception. At the highest level, the mechanism of attention pulls out objects and critical patterns from the pattern analysis subsystem to execute a visual query.

The content of visual working memory consists of "object files," to use the term of Kahneman et al., a visual spatial map in egocentric coordinates that contains residual information about a small number of recently attended objects. Also present is a visual query pattern that forms the basis for active visual search through the direction of attention.

We tend to think of objects as relatively compact entities, but objects of attention can be extended patterns, as well. For example, when we perceive a major highway on a map winding though a number of towns, that highway representation is also a visual object. So, too, is the *V* shape of a flight of geese, the pattern of notes on a musical score that characterize an arpeggio, or the spiral shape of a developing hurricane.

Following is a list of the key features of the visual thinking process:

1. Problem components are identified that have potential solutions based on visual pattern discovery. These are formulated into visual queries consisting of simple patterns.
2. Eye-movement scanning strategies are used to search the display for the query patterns.
3. Within each fixation, the query determines which patterns are pulled from the flux of pattern-analysis subsystems.
    a. Patterns and objects are formed as transitory object files from proto-object and proto-pattern space.
    b. Only a small number of objects or pattern components are retained from one fixation to the next. These object files also provide links to verbal-propositional information in verbal working memory.
    c. A small number of cognitive markers are placed in a spatial map of the problem space to hold partial solutions where necessary. Fixation and deeper processing are necessary for these markers to be constructed.
4. Links to verbal-propositional information are activated by icons or familiar patterns, bringing in other kinds of information.

Visual Problem Solving Processes

The actual process of problem solving can be represented as a set of embedded processes. They are outlined in Figure 11.12. At the highest level is problem formulation and the setting of high-level goals—this is likely to occur mostly using verbal-propositional resources.

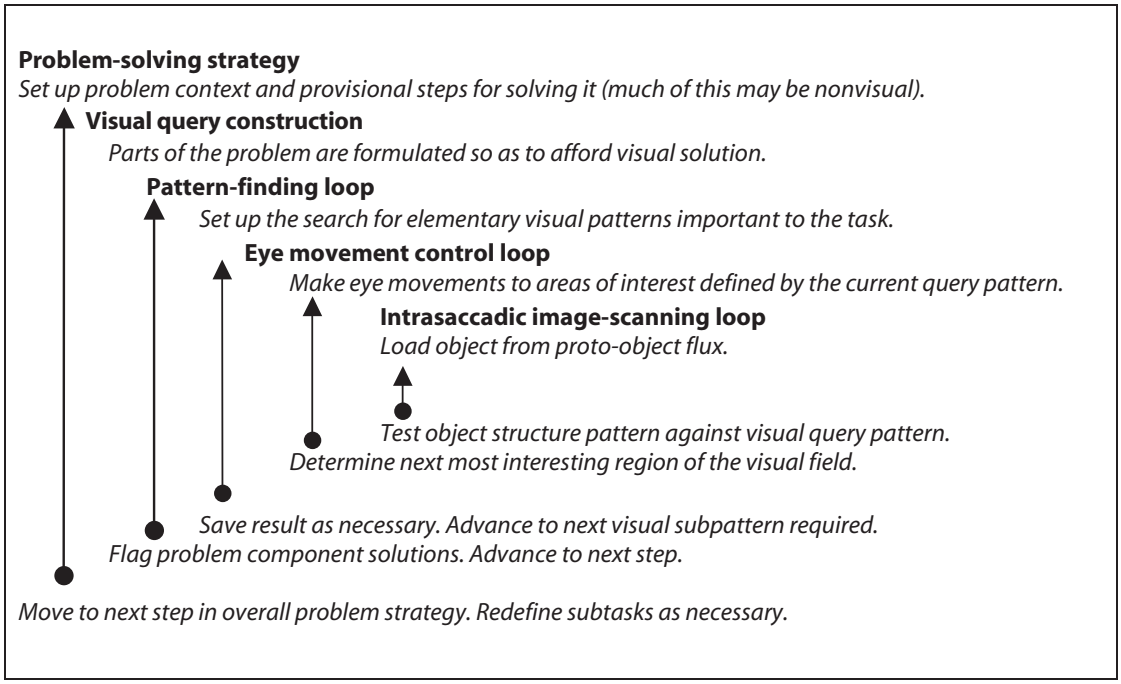

Figure 11.12 Visually aided problem solving can be considered a set of embedded processes. Interactive techniques, such as brushing, zooming, or dynamic queries, can be substituted for the eye movement control loop.

To give substance to this rather abstract description, let us consider how the model deals with a fairly common problem—planning a trip aided by a map. Suppose that we are planning a trip through France from Port-Bou in Spain, near the French border, to Calais in the northeast corner of France. The visualization that we have at our disposal is the map shown in Figure 11.13.

The Problem-Solving Strategy

The initial step in our trip planning is to formulate a set of requirements, which may be precise or quite vague. Let us suppose that for our road trip through France we have five days at our disposal and we will travel by car. We wish to stop at two or three interesting cities along the way, but we do not have strong preferences. We wish to minimize driving time, but this will be weighted by the degree of interest in different destinations. We might use the Internet as part of the process to research the attractions of various cities; such knowledge will become an important weighting factor on the alternate routes. We begin planning our route: a problem-solving strategy involving visualization.

Visual Query Construction

We establish the locations of various cities through a series of preliminary visual queries to the map. Fixating each city icon and reading its label helps to establish a connection to the verbal-

Figure 11.13 Planning a trip from Port-Bou to Calais involves finding the major routes that are not excessively long and then choosing among them. This process can be understood as a visual search for patterns.

propositional knowledge we have about that city. Little, if any, of this will be retained in working memory, but the locations will be primed for later reactivation.

Once this has been done, path planning can begin by identifying the major alternative routes between Port-Bou and Calais. The visual query we construct for this will probably not be very precise. Roughly, we are seeking to minimize driving time and maximize time at the stopover cities. Our initial query might be to find the set of alternative routes that are within 20% of the shortest route, using mostly major highways. From the map, we determine that major roads are represented as wide orange lines and incorporate this fact into the query.

The Pattern-Finding Loop

The task of the pattern-finding loop is to find all acceptable routes as defined by the previous step. The patterns to be discovered are continuous contours, mostly orange (for highways) and not overly long, connecting the start city icon with the end city icon. Two or three acceptable

solutions may be stored in visual working memory, or a single route may use the entire capacity of this limited resource.

Once a complete path has been identified, it must be retained in some way while alternate solutions are found. If a route is complex, part of the task can be held in verbal-propositional form (e.g., the western route or the Bordeaux route), and this label can be used later as the starting point for a rapid visual reconstruction. An interactive computer-based map can support this loop by allowing a user to highlight a potential solution, as illustrated for the Bordeaux, Poitiers, Paris path in Figure 11.13. This will free up more capacity in searching for alternative paths.

As a result of this process, three alternative solutions are identified. A western route goes through Toulouse, Bordeaux, Nantes, Caen, and Rouen. A central route shares a path to Bordeaux but then goes via Poitiers, Orleans, and Paris. An eastern route would get us to Paris via Montpellier, Avignon, Lyon, and Dijon.

The Eye Movement Control Loop

The detailed execution of the pattern-finding process is carried out through a series of eye movements to capture each of the major continuous paths meeting the criteria. The eye movements are planned using the task-weighted spatial map of proto-patterns. Those patterns most likely to be relevant to the current task are scheduled for attention, starting with the one weighted most significant. As part of this process, partial solutions are marked in visual working memory by setting placeholders in the egocentric spatial map. For example, the part of the route that goes to Bordeaux might be marked while the alternatives for the rest of the trip are explored. Once an entire path has been identified, it may be checked with a set of rapid eye movements.

This stage may be supported in an interactive system by some form of both spatial and semantic scaling (Furnas 1986). At the early planning stage, only major highways and good secondary roads are required, so it will be easier to carry out this task if the map is simplified to show only these. A smaller map may also be easier to parse with eye movements. Later planning stages will require more detail and zoomed-in map views.

The Intrasaccadic Scanning Loop

This is the innermost loop of the visual query system, where the information available from a single fixation is processed. Sections of lines representing roads are successively formed through selective tuning of the pattern-finding mechanism (Dickinson et al., 1997). Those representing minor roads going in the wrong direction will be rejected, whereas those representing connected major roads going in the right direction will be held in visual working memory up to a limit of three or four road segments. City names will also be processed, causing information about them to be loaded into verbal-propositional memory.

Implications for Interactive Visualization Design

The model presented here has a number of implications for data display systems. The following are perhaps the three most important:

1. Data should be presented in such a way that informative patterns are easy to perceive, allowing us to take advantage of the pattern-finding capabilities of the middle stage of visual processing.
2. The cognitive impact of the interface should be minimized, so that thinking is about the problem, not the interface.
3. The interface should be optimized for low-cost, rapid information seeking.

In the following sections, we consider these implications in more detail, first speculating on the nature of visual queries and then in the context of interactive data navigation techniques, both in isolation and as they are encapsulated in various experimental applications.

Visual Query Patterns

The patterns that can be part of visual problem solving are infinitely diverse: path finding in graphs, quantity estimation, magnitude estimation, trend estimation, cluster identification, correlation identification, outlier detection and characterization, target detection, identification of structural patterns (e.g., hierarchy, degree of coupling), to name a few.

For a visual query to be performed rapidly and with a low error rate, it should consist of a simple pattern or object that can be held in visual working memory. Three elementary queries, or one more complex query, can be held in visual working memory. Other cognitive strategies are required when a query is more than the capacity of visual working memory. Figure 11.14

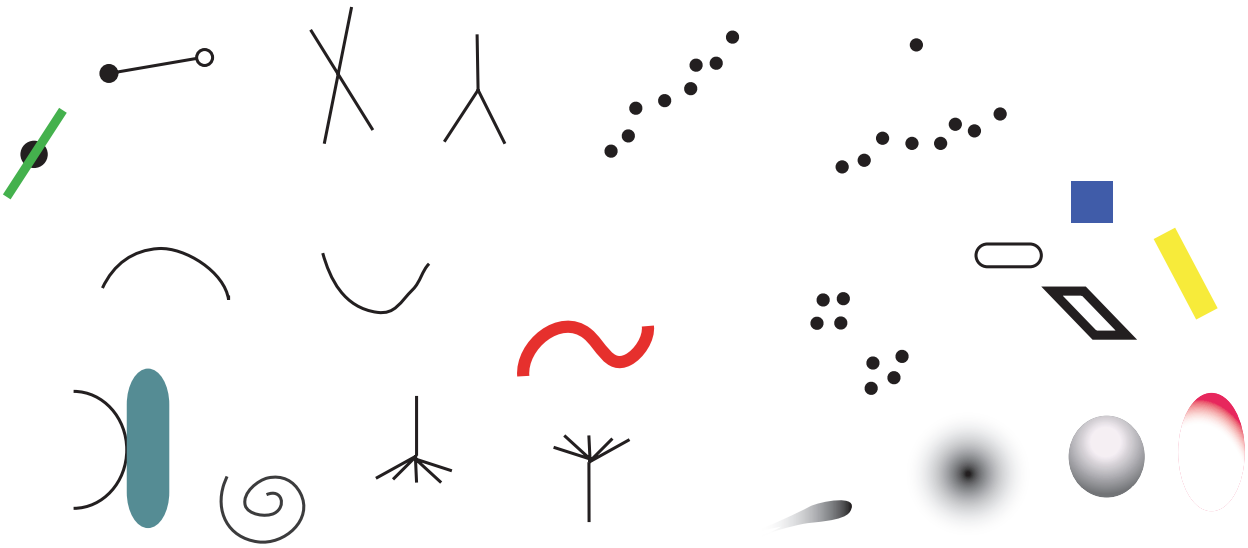

Figure 11.14 Any simple pattern can form the basis for a visual query. Expertise with a particular kind of visual display will allow for more complex queries.

gives some suggestions of patterns that might constitute simple queries. The number of possible patterns is astronomical. Knowledge from vision research about pattern perception and preattentive vision can provide a good understanding of the kinds of visual queries that can be processed rapidly. After all, most studies of perception take the form of having subjects make repeated visual queries of a display.

We may be able to query patterns of considerably greater complexity as we become expert in a particular set of graphical conventions, such as a circuit diagram. A chess master can presumably make visual queries consisting of patterns that would not be possible for a novice. Nevertheless, even for the expert, the laws of preattentive processing and elementary pattern perception will make certain patterns much easier to see than others.

Costs of Navigation

In the previous chapter, we discussed various methods for navigating through information spaces. Now we briefly reconsider these in the light of the current theory. The benchmark point of comparison for all navigation techniques should be saccadic eye movement. This allows us to acquire a new set of informative visual objects in 100–200 msec. Moreover, information acquired in this way will be integrated readily with other information that we have recently acquired from the same space. Thus, the ideal visualization is one in which all the information for visualization is available on a single high-resolution screen. Even if the information is in different windows, the cost of navigating there is only a single eye movement, or in the worst case, an eye movement plus a head movement if the angle is large. Consider some of the computer-based alternatives to eye movements.

Hypertext Link

Clicking a hypertext link involves a 1–2 sec guided hand movement and a mouse click. This can generate an entirely new screenful of information. However, the cognitive cost is that the entire information context typically has changed, and the new information may be presented using a different visual symbol set and different layout conventions. Several seconds of cognitive reorientation may be required.

Brushing, Dynamic Queries, and Hover Queries

Both brushing (Ahlberg et al., 1992) and dynamic queries (Becker and Cleaveland, 1987) allow information to be revealed on some data dimension by making a continuous mouse movement. Hover queries cause extra information to pop up rapidly as the mouse is dragged over a series of data objects. All three of these require a mouse movement typically taking about two seconds. After this initial setup time, the mouse can allow rapid scanning in a tight, exploratory visual feedback. The data is continuously modified according to the mouse movement. This may enable an effective query rate of several per second, similar to the rate for eye movements. However, this rate is only possible for quite specific kinds of query trajectory; we cannot jump from point to point in the data space as we can by moving our eyes.

Walking or Flying in Virtual Reality

Compared to eye movements or rapid exploration techniques like hyperlink following or brushing, navigating a virtual information space by walking or flying is likely to be both considerably slower and cognitively more demanding. In virtual reality, as in the real world, walking times are measured in minutes at best. Even with virtual flying interfaces (which do not attempt to simulate real flying and are therefore much faster), it is likely to take tens of seconds to navigate from one vantage point to another. In addition, the cognitive cost of manipulating the flying interface is likely to be high without extensive training. And although walking in virtual reality simulates walking in the world, it cannot be the same, so the cognitive load is higher.

Table 11.1 gives a set of rough estimates of the times and cognitive costs associated with different navigation techniques. When simple pattern finding is needed, the importance of having a fast, highly interactive interface cannot be emphasized enough. If a navigation technique is slow, then the cognitive costs can be much greater than just the amount of time lost, because an entire train of thought can become disrupted by the loss of the contents of both visual and nonvisual working memories.

The figures in Table 11.1 should be taken as ballpark estimates; they have not been empirically validated.

Magnifying Windows vs. Zooming

As an example of how visual working memory capacity can be used to make substantial design decisions, we now consider the problem of when extra windows are needed in a visualization interface. Consider the task of finding similar or identical patterns spaced far apart in a large geographical space, as illustrated in Figure 11.15. With a zooming interface, it is necessary to

Navigation Technique	Time	Cognitive Effort
Attentional object switch within a fixation	50 msec	Minimal
Saccadic eye movement	150 msec	Minimal
Hypertext jump	2 sec	Medium
Brushing	2 sec setup, 250 sec/query	Medium
Dynamic queries	2 sec setup, 250 sec/query	Medium
Floating queries	2 sec setup, 250 sec/query	Medium
Zooming	2 sec setup + log spatial change	High
Flying	30 sec–3 min	High
Walking	30 sec–10 min	High

Table 11.1 Approximated times for navigation in information spaces

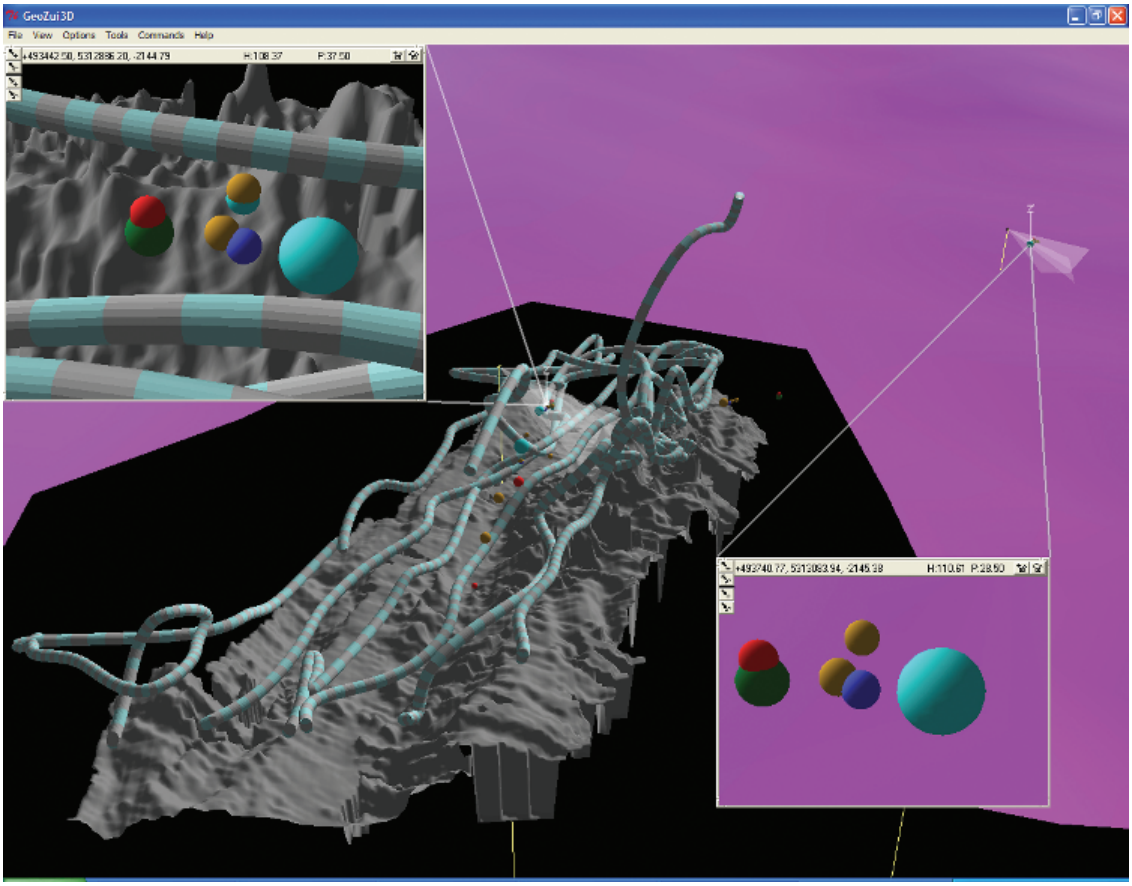

Figure 11.15 Subwindows show a magnified view together with the source of the information in the background overview.

zoom in to and look at one pattern, then hold that pattern in visual working memory while zooming out to seek other patterns. The pattern in visual working memory is then compared to new patterns seen during the search process. If a possible match is found, it may be necessary to zoom back and forth to confirm details of the match. An alternative method is to use extra windows to magnify parts of the main display. When two such windows are in position, it is possible simply to make eye movements between them to assess the match more rapidly.

The critical resource here is visual working memory capacity, because this determines how many visits are required to make the comparison. If the target pattern is simple enough to be held in visual working memory, then zooming will often be more efficient, because it avoids the overhead of setting up multiple windows. If more than three items are in the target pattern, then it will be necessary to zoom back and forth between them, and the multiwindow solution will be faster.

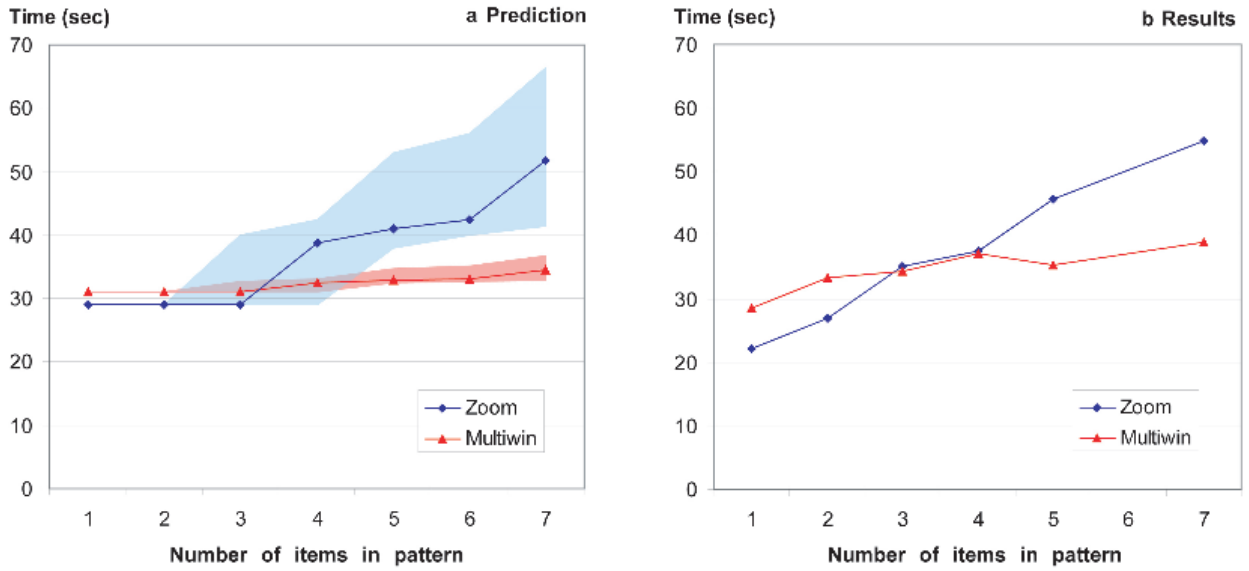

Figure 11.16 Model predictions are shown on the left. Measured task performance is shown on the right. Multiple windows speed performance relative to the use of a zooming interface when the number of objects to be compared is five or more.

We have modeled two user interfaces—simple zooming vs. multiple windows—considering visual working memory capacity as a critical resource and estimating the number of zooms vs. window movements necessary to complete the task of finding identical groups of simple shapes clustered but widely separated in a geographical space (Plumlee and Ware, 2002). The predictions of the model are shown in Figure 11.16(a), modeled for capacities of visual working memory at two, three, and four items, leading to a range of predictions as shown by the broad colored wedges. As can be seen, the model predicts that zooming will have an initial benefit because extra windows take more time to set up. However, as the number of objects increases, the extra window interface will be beneficial. The measured results, as shown Figure 11.16(b), closely matched the prediction.

Interfaces to Knowledge Structures

Although a visual icon or object can activate logical verbal information rapidly and effectively, it can do so only if this information has a strong, previously learned association between the image and the meaning. This is why advertisers spend millions promoting logos. It is also why the design of symbol sets is so critical. Once a symbol set has been learned by a wide group of users, the cost of changing it can be huge.

Concept Maps and Mind Maps

A more complex way to map out knowledge structures is to use some form of node–link diagram, with nodes representing concepts and the links representing the relationships between them. The

technique of sketching out links between concepts, as shown in Figure 11.10, has received considerable attention from educational theorists. These *concept maps* (or *mind maps* as they are sometimes called) are often recommended as study aids for students (Jonassen et al., 1993). Usually, such maps are constructed informally by sketching them on paper, but computer-based tools also exist. Essentially, a concept map is a type of node–link diagram in which the nodes represent concepts and the links represent relationships between concepts. It can be used to make the structure of a cognitive concept network explicitly available. An individual can use a concept map as a tool for organizing his or her own personal concept structure, and it may reveal patterns of relationships between ideas that had not been evident when the concepts were stored internally. A concept map can also be constructed in a group exercise, in which case it becomes a tool for building a common understanding.

Constellation

The Constellation system of Munzner et al. (1999) provides an example of how a highly interactive node–link visualization can provide views into very a complex semantic network, far larger than can be displayed on a static concept map. Figure 11.17 shows a screen shot, but this static image does not do justice to the system. Constellation uses hover queries to allow for rapid highlighting of subsets of the graph. Links attached to a node became highlighted as the cursor passes over the node. In addition, when the user clicks on a particular node, Constellation uses intelligent zoom, causing the graph to rearrange itself partially so that closely related semantic concepts are allocated more screen space and larger fonts. By using these techniques, a large amount of semantic information can be accessed very quickly.

Note that the rapid query techniques get around the usual problems of graph layout. Most of the work in graph layout is aimed at producing aesthetically pleasing drawings of graph structures, by paying particular attention to minimizing edge crossings of nodes (Di Battista et al., 1998). A good, clear static graph drawing of the information in Figure 11.17 by the conventional criteria, is probably impossible, because there are simply too many links. In Constellation, Munzer abandoned the usual criteria, allowing edges to cross each other and to cross nodes. Using interactive techniques to reveal information as needed allowed visual access to much larger structures.

The node–link diagram is a method for looking at networks of concepts, but a common way of organizing knowledge is through a hierarchy, and the most common visualization of a hierarchy is a tree. A degree-of-interest tree (Card and Nation, 2002) is a tree visualization that uses a *degree-of-interest function* (Furnas, 1986) to show or hide interactively parts of the tree structure based on their estimated relevance to a selected node. It enables a large tree structure to be interactively explored.

Linking Computer-Based Analysis with Visualization

The greatest power in information visualization arises when the power of computers to sample and condense very large amounts of information is combined with a visual interface. If the computer contains a model of a knowledge domain, then this model can form the basis for inter-

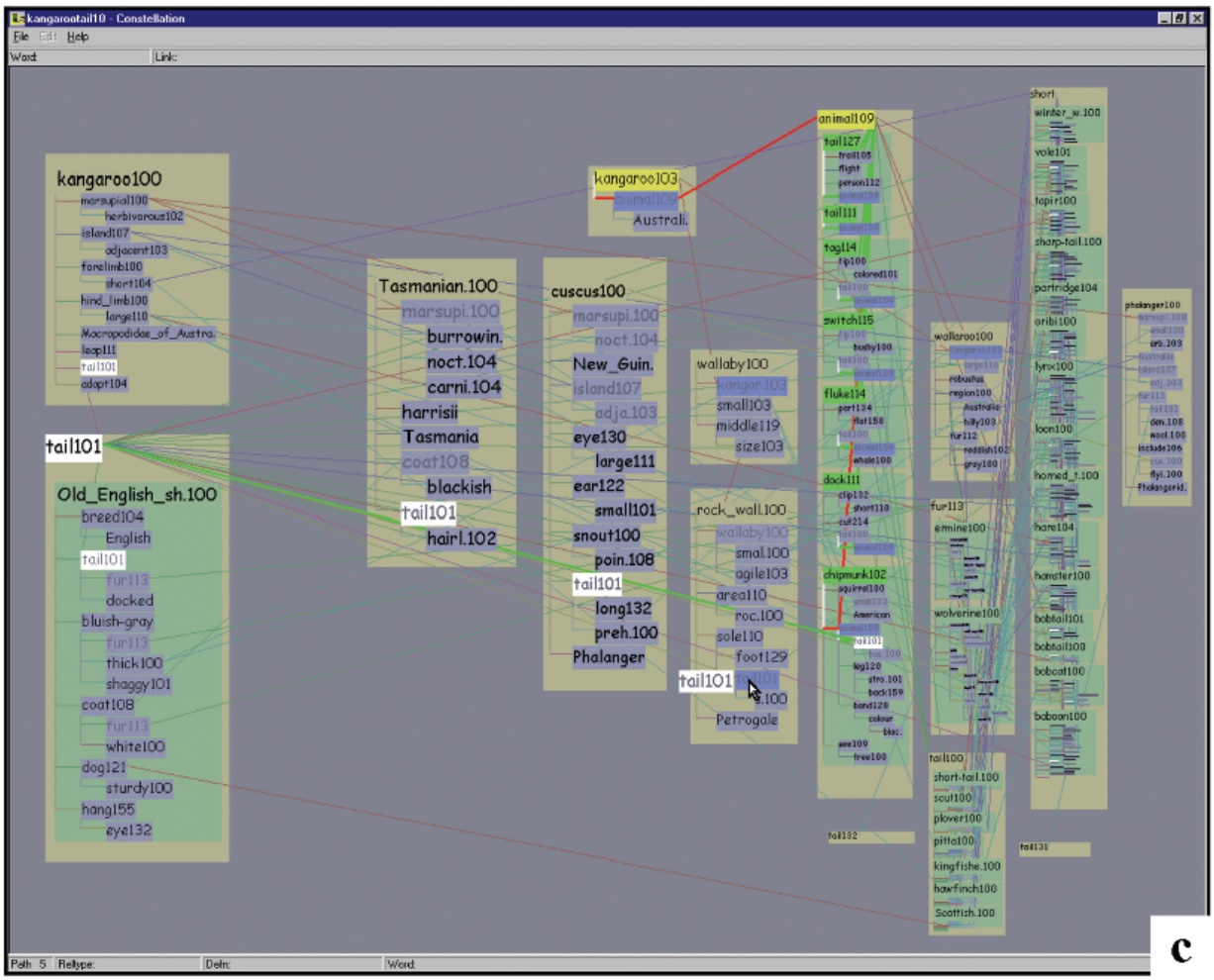

Figure 11.17 A screen image of the Constellation system showing a view into the MindNet semantic network database.

active applications that visually hide and show information based on the estimated relevance to some task. Only the most relevant information is being displayed to the user.

Psychologists have developed a number of other tools for mapping the cognitive structure of concepts, besides simple sketching. One is based on multidimensional scaling (Shepard, 1962). The technique involves giving participants pairs of examples of the ideas or objects to be mapped and asking them to rate the similarity. For example, if the goal is to find out how someone conceptualizes different kinds of animals, that person is given pairs of cat–dog, mouse–cow, cat–elephant, and so on, and asked to give each pair a similarity rating. Once all pair ratings for the entire set have been obtained, the multidimensional scaling technique is used to compute a mathematical space in which similar animals are close together. This technique can sometimes reveal

the nature of the most significant dimensions of this space, but often the mathematical dimensions that are found have little intuitive meaning. The multidimensional scaling technique does not show links between concepts; it only shows proximity. Concepts that are close together in the space are assumed to be related.

Multidimensional scaling can be used as a tool in visualizing concept spaces, but it suffers from the problem that the space created can have a high dimensionality. However, the dimensionality can be reduced by simply showing the two or three most significant dimensions as a 2D or 3D scatter plot. More dimensions can be added by color-coding or changing the shape of each data glyph, as discussed in Chapters 4 and 5.

The analysis of large text databases is an application area where it is useful to get a view of a large number of points in a multidimensional conceptual space. The SPIRE system creates a classification of documents with respect to a keyword query and can be applied to databases consisting of hundreds of thousands of documents (Wise et al., 1995). The result is a set of vectors in an *n*-dimensional space. To help people understand the resulting clusters of documents, Wise et al. created a visualization called a *ThemeScape*, which shows the two most important dimensions as a kind of data landscape. This is illustrated in Figure 11.18. Flags on the tops of hills

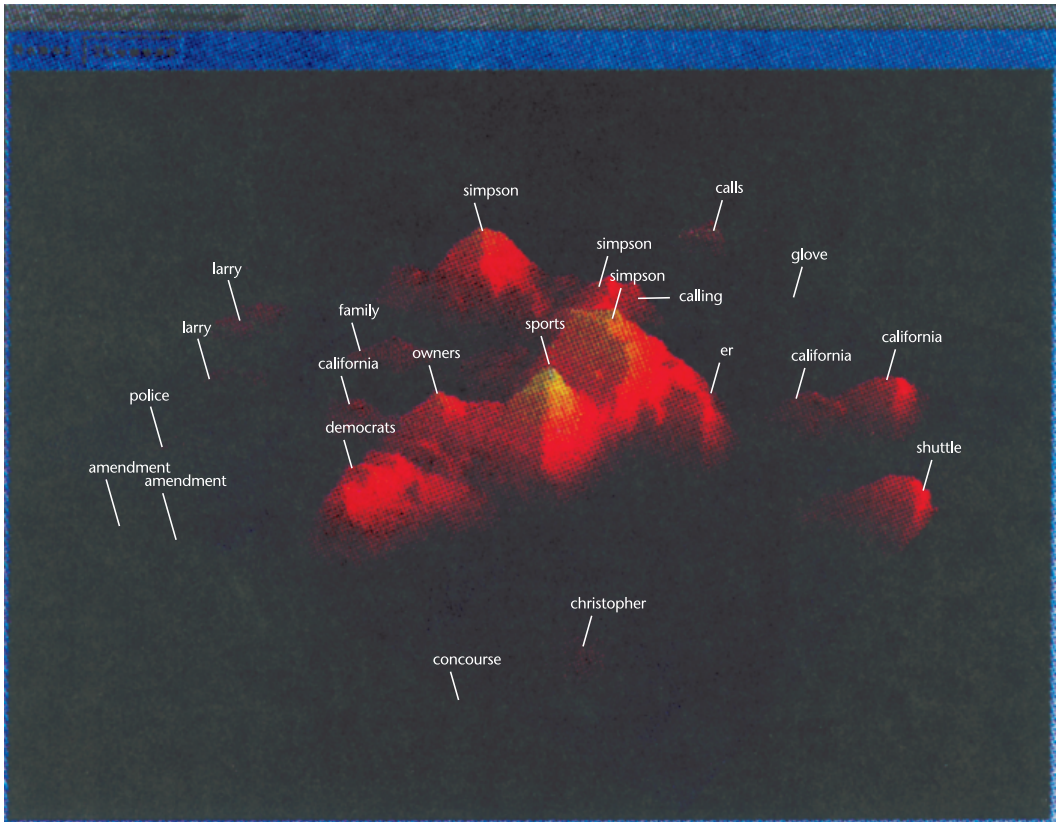

Figure 11.18 An entire week of CNN news stories is summarized in a ThemeScape visualization (Wise et al., 1995).

label and identify the largest clusters of documents in this space. Essentially, a ThemeScape uses the two most significant dimensions of the space to create a smoothed two-dimensional histogram. This can be regarded as a different kind of concept map—one that does not show the links but uses spatial proximity and salience to show the major concentrations of information and, to some extent, their relationships. This kind of display will be useful when two dimensions really do capture most of the variability in the data. If more dimensions are significantly involved, then color-coding and more interactive exploratory techniques may be necessary.

Trajectory Mapping

Trajectory mapping is a recent psychological method for mapping out the structure of concept spaces (Richards and Koenderink, 1995). Unlike multidimensional scaling, trajectory mapping explicitly finds links between concepts. In trajectory mapping, a participant is also given pairs of examples from the set of objects (or concepts) to be organized. However, in this case the person is asked to look at the objects that make up the pair and extrapolate on the basis of some difference between them, then select another sample concept that represents the result of that extrapolation. For example, someone given a mouse and a dog as exemplars might extrapolate to a cow if they thought size was a critical variable, or might extrapolate to a monkey based on a concept of animal intelligence. Participants are also allowed to say that there is no meaningful extrapolation, in which case one of the exemplars becomes a terminator in the resulting concept graph. This exercise is designed to produce a set of cognitive pathways linking concepts. Strong pathways can be distinguished from weak ones.

Lokuge et al. (1996) used a combination of trajectory mapping and multidimensional scaling to create different visual maps linking various tourist attractions in the Boston area, such as museums and open-air markets. The results were based both on conceptual similarity between the different items and the pathways between them. One of the results is shown in Figure 11.19. This technique could be used to generate customized tours automatically. The tourist would enter a set of interests, and the system would combine this with the database information to create a walking tour of suitable attractions.

It should be recognized that no matter how they are generated, concept maps are somewhat crude instruments for making knowledge explicit. All of them reveal only that there is a relationship between ideas, not the nature of that relationship.

Creative Problem Solving

We commonly divide problem-solving activities into the routine and the creative. The essential difference is that in creative thinking, the emphasis is on novelty. Theories of creative thinking generally break the process into three states: preparation, production, and judgment (Matlin, 1994). Visualization can help with all three.

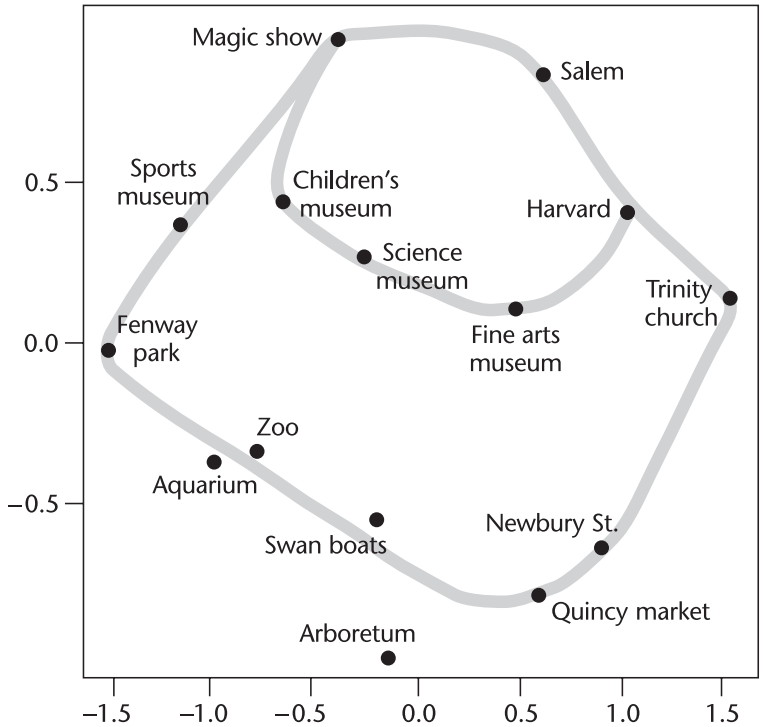

Figure 11.19 A trajectory map of tourist destinations in the Boston area, laid out according to the results of a multidimensional scaling experiment (Lokuge et al., 1996).

In the *preparation* stage, the problem-solver acquires the background information needed to build a solution. Sometimes, preparation will involve a stage of exploratory data analysis. Visualization can help through pattern discovery (discussed especially in Chapter 6). In this process, the visual queries may initially be a loosely defined search for any significant pattern, becoming more focused as the issues become better defined.

In the *production* stage, the problem-solver generates a set of potential solutions. A solution often starts with a tentative suggestion, which is either rejected or later refined. Early theorists proposed that the quantity of ideas, rather than the quality, was the overriding consideration in the production of candidate solutions. However, experimental studies fail to support this (Gilhooly, 1988). Generating ideas irrespective of their value is probably not useful.

Possibly the most challenging problem posed in data visualization systems is to support the way sketchy diagrams are used by scientists and engineers in the production stage. Discoveries and inventions made using table-napkin sketches are legendary. Here is a description of the role of a diagram by an architectural theorist (Alexander, 1964):

> *Each constructive diagram is a tentative assumption about the nature of the context. Like a hypothesis, it relates an unclear set of forces to one another conceptually; like a hypothesis, it is usually improved by clarity and economy of notation. Like a hypothesis,*

> *it cannot be obtained by deductive methods, but only by abstraction and invention. Like a hypothesis, it is rejected when a discrepancy turns up and shows that it fails to account for some new force in the context.*

If creativity is to be supported, the medium must afford tentative interactions. The lack of precision in quick, loose sketches actually allows for multiple interpretations. The fact that a line can be interpreted in many different ways, as discussed in Chapter 6, can be a distinct benefit in enabling a diagram to support multiple tentative hypotheses. The sketches people construct as part of the creative process are rapid, not refined, and readily discarded. Giving a child high-quality watercolor paper and paints is likely to inhibit creativity if the child is made aware of the expense and cautioned not to "waste" the materials. Schumann et al. (1996) carried out an empirical study of architectural perspective drawings executed in three different styles: a precise line drawing, a realistically shaded image, and a sketch. All the drawings contained the same features and level of detail. The sketch version was rated substantially higher on measures of ability to stimulate creativity, changes in design, and discussions.

In the *judgment* stage, the problem-solver analyzes the potential solutions. This stage is an exercise in quality control; as fast as hypotheses are created and patterns are discovered, most must be rejected. In a visualization system used for data mining, the user may discover large numbers of patterns but will also be willing to reject them almost as rapidly as they are discovered. Some will already be known, some will be irrelevant to the task at hand, only a few will be novel, and even fewer will lead to a practical solution. Many judgment aids are not visual; for example, statistical tools can be used to test hypotheses formally. But when visualization is part of the process, it should not be misleading or hide important information.

The challenge for problem-solving interfaces is to support the rapid creation of loose sketches, the ability to modify them, and the ability to discard all or some of them. All this must be done with an interface so simple that it does not intrude on the visual thinking process.

Conclusion

The best visualizations are not static images to be printed in books, but fluid, dynamic artifacts that respond to the need for a different view or for more detailed information. In some cases, the visualization can be an interface to a simulation of a complex system; the visualization, combined with the simulation, can create a powerful cognitive augmentation. An emerging view of human–computer interaction considers the human and the computer together as a problem-solving system. The visualization is a two-way interface, although highly asymmetric, with far higher bandwidth communication from the machine to the human than in the other direction. Because of this asymmetry in data rates, cognitive support systems must be constructed that are semiautomatic, with only occasional nudges required from users to steer them in a desired direction. The high-bandwidth visualization channel is then used to deliver the results of modeling exercises and database searches.

At the interface, the distinction between input and output becomes blurred. We are accustomed to regarding a display screen as a passive output device and a mouse as an input device. This is not how it is in the real world, where many things work both ways. A sheet of paper or a piece of clay can both record ideas (input) and display them (output). The coupling of input and output can also be achieved in interactive visualization. Each visual object in an interactive application can potentially provide output as a representation of data and also potentially receive input. Someone may click on it with a mouse to retrieve information or may use it as an interface to change the parameters of a computer model. The ultimate challenge for this kind of highly interactive information visualization is to create an interface that supports creative sketching of ideas, affording interactive sketching that is as fluid and inconsequential as the proverbial paper-napkin sketch.

The person who wishes to design a visualization must contend with two sets of conflicting forces. On the one hand, there is the requirement for the best possible visual representation, tailored exactly to the problem to be solved. On the other hand, there is the need for consistency in representation any time that two or more people work on a problem. This need is even greater when large, international organizations have a common set of goals that demand industrywide visualization standards. At the stage of new discoveries in information visualizations, standardization is the enemy of innovation and innovation is the enemy of standardization. Thus it is important to get the research done before the standards are formed, otherwise it will be too late.

These are exciting times for information visualization, because we are still in the discovery phase, although this phase will not last for long. In the next few years, the wild inventions that are now being implemented will become standardized. Like clay sculptures that have been baked and hardened, the novel data visualization systems of today's laboratory will become cultural artifacts, everyday tools of the information professional.