CHAPTER 7

Visual Objects and Data Objects

The object metaphor is pervasive in the way we think about abstract data. Object-oriented programming is one example; the body politic is another. Object-related concepts are also basic in modern systems design. A modular system is one that has easily understood and easily replaced components. Good modules are *plug-compatible* with one another; they are discrete and separate parts of a system. In short, the concept of a module has a lot in common with the perceptual and cognitive structures that define visual objects. This suggests that visual objects may be an excellent way to represent modular system components. A visual object provides a useful metaphor for encapsulation and cohesiveness, both important concepts in defining modular systems.

For our present purposes, an *object* can be thought of as any identifiable, separate, and distinct part of the visual world. Information about visual objects is cognitively stored in a way that ties together critical features, such as oriented edges and patches of color and texture, so that they can be identified, visually tracked, and remembered. Because visual objects cognitively group visual attributes, if we can represent data values as visual features and group these features into visual objects, we will have a very powerful tool for organizing related data.

Two radically different theories have been proposed to explain object recognition. The first is image-based. It proposes that we recognize an object by matching the visual image with something roughly like a snapshot stored in memory. The second type of theory is structure-based. It proposes that is analyzed in terms of primitive 3D forms and the structural interrelationships between them. Both of these models have much to recommend them, and it is entirely plausible that each is correct in some form. It is certainly clear that the brain has multiple ways of analyzing visual input. Certainly, both models provide interesting insights into how to display data effectively.

Image-Based Object Recognition

We begin with some evidence related to picture and image perception. People have a truly remarkable ability to recall pictorial images. In an arduous experiment, Standing et al. (1970) presented subjects with a list of 2560 pictures at a rate of one every 10 seconds. This was like the family slide show from hell, it took them more than seven hours spread over a four-day period. Amazingly, when subsequently tested, subjects were able to distinguish pictures from others not previously seen, with better than 90% accuracy.

People can also recognize objects in images that are presented very rapidly. Suppose you asked someone, "Is there a dog in one of the following pictures?" and then showed them a set of images, rapidly, all in the same place, at a rate of 10 per second. Remarkably, they will be able to detect the presence, or absence, of a dog in one of the images most of the time. This experimental technique is called *rapid serial visual presentation* (RSVP). Experiments have shown that the maximum rate for the ability to detect common objects in images is about 10 images per second (Potter and Levy, 1969; Potter, 1976).

A related phenomenon is *attentional blink*. If, in a series of images, a second dog were to appear in an image within 350 ms of the first, people do not notice it (or anything else). This moment of blindness is the attentional blink (Coltheart, 1999). It is conjectured that the brain is still processing the first dog, even though the image is gone, and this prohibits the identification of other objects in the sequence.

It is useful to make a distinction between recognition and recall. We have a great ability to recognize information that we have encountered before, as the picture memory experiment of Standing et al. shows. However, if we are asked to reconstruct visual scenes—for example, to recall what happened at a crime scene—our performance is much worse. Recognition is much better than recall. This suggests that a major use of visual images can be as an aid to memory. An image that we recognize can help us remember events or other information related to that image. This is why icons are so effective in user interfaces; they help us to recall the functionality of computer programs.

More support for image-based theories comes from studies showing that three-dimensional objects are recognized most readily if they are encountered from the same view direction as when they were initially seen. Johnson (2001) studied subjects' abilities to recognize bent pipe structures. Subjects performed well if the same viewing direction was used in the initial viewing and in the test phase; they performed poorly if a different view direction was used in the test phase. But subjects were also quite good at identification from exactly the opposite view direction. Johnson attributed this unexpected finding to the importance of silhouette information. Silhouettes would have been similar, although flipped left-to-right from the initial view.

Although most objects can easily be recognized independent of the size of the image on the retina, image size does have some effect. Figure 7.1 illustrates this. When the picture is seen from a distance, the image of the Mona Lisa face dominates; when it is viewed up close, smaller objects become dominant: a gremlin, a bird, and a claw emerge. Experimental work by Biederman and

Figure 7.1 When the image of the Mona Lisa is viewed from a distance, the face dominates. But look at it from 30 cm, and the gremlin hiding in the shadows of the mouth and nose emerges. When component objects have a size of about 4 degrees of visual angle, they become maximally visible. *Adapted from the work of the Tel Aviv artist Victor Molev.*

Cooper (1992) suggests that the optimal size for recognizing a visual object is about 4 to 6 degrees of visual angle. This gives a useful rule of thumb for the optimal size for rapid presentation of visual images so that we can best see the visual patterns contained in them.

Another source of evidence for image-based object recognition comes from priming effects. The term *priming* refers to the fact that people can identify objects more easily if they are given prior exposure to some relevant information. Most priming studies have been carried out using verbal information, but Kroll and Potter (1984) showed that *pictures* of related objects, such as a cow and a horse, have a mutually priming effect. This is similar to the priming effect between the words *cow* and *horse*. However, they found little cross-modality priming; the word *cow* provided only weak priming for a picture of a horse. It is also possible to prime using purely visual information, that is, information with no semantic relationship. Lawson et al. (1994) devised a series of experiments in which subjects were required to identify a specified object in a series of

briefly presented pictures. Recognition was much easier if subjects had been primed by visually similar images. They argued that this should not be the case if objects are recognized on the basis of a high-level, 3D structural model of the kind that we will discuss later in this chapter; only image-based storage can account for their results.

Priming effects can occur even if information is not consciously perceived. Bar and Biederman (1998) showed pictorial images to subjects, so briefly that it was impossible for them to identify the objects. They used what is called a *masking technique*, a random pattern shown immediately after the target stimulus to remove the target from the iconic store, and they rigorously tested to show that subjects performed at chance levels when reporting what they had seen. Nevertheless, 15 minutes later, this unperceived exposure substantially increased the chance of recognition on subsequent presentation. Although the information was not consciously perceived, exposure to the particular combination of image features apparently primed the visual system to make subsequent recognition easier. They found that the priming effect decreased substantially if the imagery was displaced sideways. They concluded that the mechanism of priming is highly image-dependent and not based on high-level semantic information.

Palmer et al. (1981) showed that not all views of an object are equally easy to recognize. They found that many different objects have something like a *canonical view* from which they are most easily identified. From this and other evidence, a theory of object recognition has been developed, proposing that we recognize objects by matching the visual information with internally stored viewpoint-specific exemplars, or "prototypes" (Edelman and Buelthoff, 1992; Edelman, 1995). According to this theory, the brain stores a number of key views of objects. These views are not simple snapshots; they allow recognition despite simple geometric distortions of the image that occur in perspective transformation. This explains why object perception survives the kinds of geometric distortions that occur when a picture is viewed and tilted with respect to the observer. However, there are strict limits on the extent to which we can change an image before recognition problems occur. For example, numerous studies show that face recognition is considerably impaired if the faces are shown upside down (Rhodes, 1995).

Adding support to the multiple-view, image-based theory of object recognition is neurophysiological data from recordings of single cells in the inferotemporal cortexes of monkeys. Perrett et al. (1991) discovered cells that respond preferentially to particular views of faces. Figure 7.2 shows some of their results. One cell (or cell assembly) responds best to a three-quarter view of a face; another, to profiles, either left or right; still another responds to a view of a head from any angle. We can imagine a kind of hierarchical structure, with the cell assemblies that respond to particular views feeding into higher-level cell assemblies that respond to any view of the object.

Applications of Images in User Interfaces

The fact that visual images are easily recognized after so little exposure suggests that icons in user interfaces should make excellent memory aids, helping us recall the functionality of parts of complex systems. Icons that are readily recognized may trigger activation of related concepts

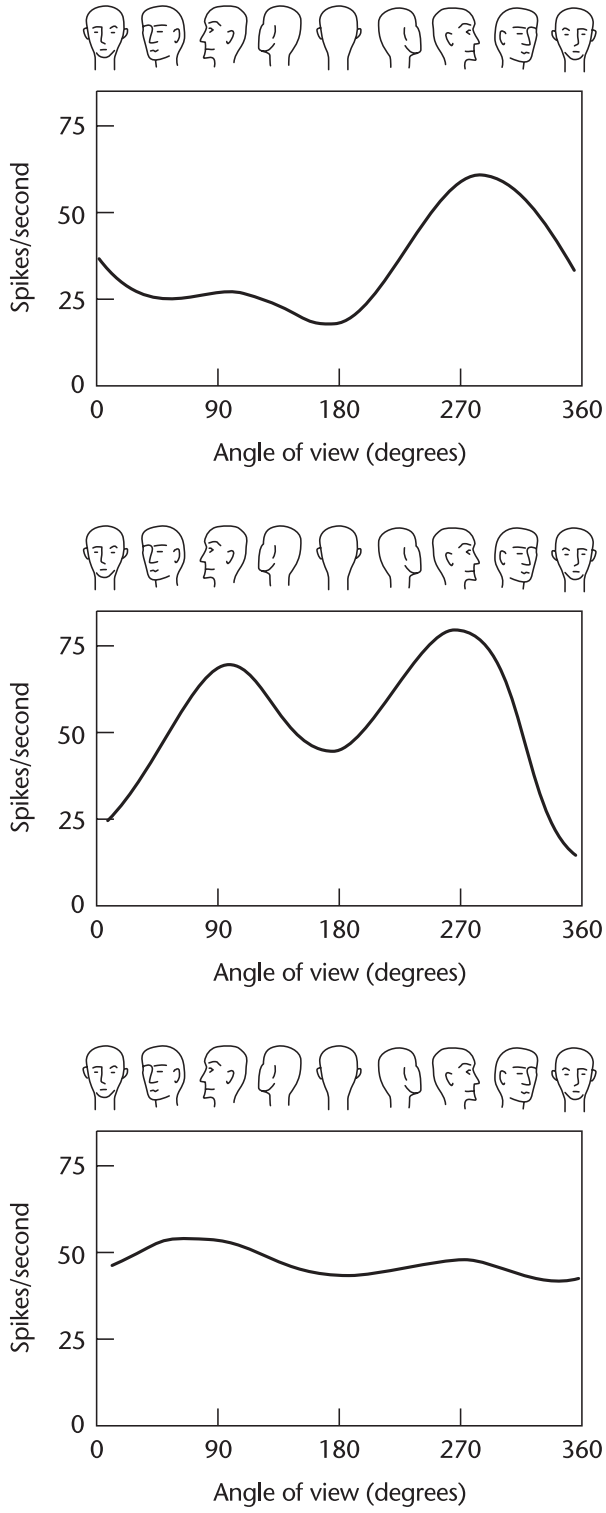

Figure 7.2 The responses of three cells in the temporal cortex of a monkey to faces in different orientations. At the top is a cell most sensitive to a right profile. The middle cell responds well to either profile. The cell at the bottom responds well to a face irrespective of orientation. *Adapted from Perrett et al. (1991).*

in the semantic network of long-term memory. Icons are also helpful because to some extent they can represent pictorially the things they are used to reference.

Priming may be useful in helping people search for particular patterns in data. The obvious way of doing this is to provide sample images of the kind of pattern being sought and repeating the samples at frequent intervals during the search process. An example would be the use of images of sample viruses in a medical screening laboratory.

Searching an Image Database

Presenting images rapidly in sequence may be a useful way to allow users to scan picture databases (Wittenburg et al., 1998; de Bruijn et al., 2000). The fact that people can search rapidly for an image in a sequence of up to 10 pictures per second suggests that presenting images using RSVP may be efficient. Contrast this with the usual method of presenting image collections in a regular grid of small thumbnail images. If it is necessary to make an eye movement to fixate each thumbnail image, it will not be possible to scan more than three to four images per second.

Even though RSVP is promising, there are a number of design problems that must be solved in building a practical interface. Once a likely candidate image is identified as being present in an RSVP sequence, it must still be found. By the time a user responds with a mouse click several images will have passed, more if the user is not poised to press the stop button. Thus, either controls must be added for backing up through the sequence, or part of the sequence must be fanned out in a conventional thumbnail array to confirm that candidate's presence and study it further (Spence, 2002; Wittenburg et al., 1998).

Personal Image Memory Banks

Based on straightforward predictions about the declining cost and increasing capacity of computer memory, it will soon be possible to have a personal memory data bank containing video and sound data collected during every waking moment of a person's lifetime. This could be achieved with an unobtrusive miniature camera, perhaps embedded in a pair of eyeglasses, and assuming continuing progress in solid-state storage, the data could be stored in a device weighing a few ounces and costing a few hundred dollars. Storing speech information will be even more straightforward. The implications of such devices are staggering. Among other things, it would be the ultimate memory aid—the user would never have to forget anything. However, a personal visual memory device of this kind would need a good user interface. One way of searching the visual content might be by viewing a rapidly presented sequence of selected frames from the video sequence. Perhaps 100 per day would be sufficient to jog the user's memory about basic events. Video data compressed in this way might make it possible to review a day in a few seconds, and a month in a few minutes.

Structure-Based Object Recognition

Image-based theories of object recognition imply a rather superficial level of analysis of visual objects. However, there is evidence that a much deeper kind of structural analysis must also occur. Figure 7.3 shows two novel objects, probably never seen by the reader before. Yet despite the fact that the *images* of these two objects are very different from one another, they can be rapidly recognized as representations of the same object. No image-based theory can account for this result.

Geon Theory

Figure 7.4 provides a somewhat simplified overview of a neural-network model of structural object perception, developed by Hummel and Biederman (1992). This theory proposes a hierarchical set of processing stages leading to object recognition. Visual information is decomposed first into edges, then into component axes, oriented blobs, and vertices. At the next layer, three-dimensional primitives such as cones, cylinders, and boxes, called *geons*, are identified. A selection of geons is illustrated in Figure 7.5. Next, the structure is extracted that specifies how the geon components interconnect; for example, in a human figure, the arm cylinder is attached near the top of the torso box. Finally, object recognition is achieved.

Silhouettes

Silhouettes appear to be especially important in determining how we perceive the structure of objects. The fact that simplified line drawings are often silhouettes may, in part, account for our ability to interpret them. At some level of perceptual processing, the silhouette boundaries of objects and the simplified line drawings of those objects excite the same neural contour-extraction mechanisms. Halverston (1992) noted that modern children tend to draw objects on the basis of the most salient silhouettes, as did early cave artists. Many objects have particular

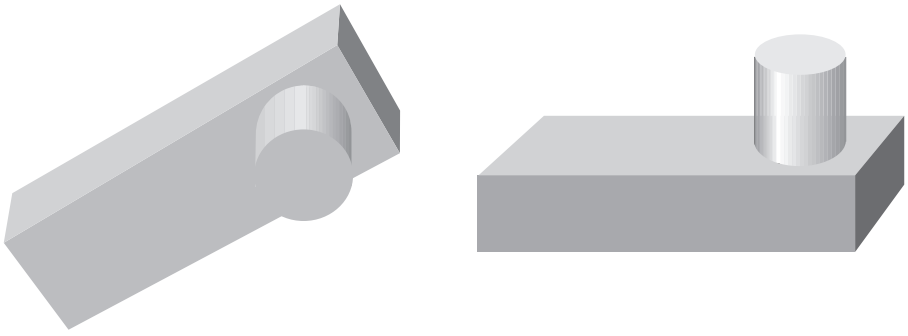

Figure 7.3 These two objects are rapidly recognized as identical, or at least very similar, despite the very different visual images they present.

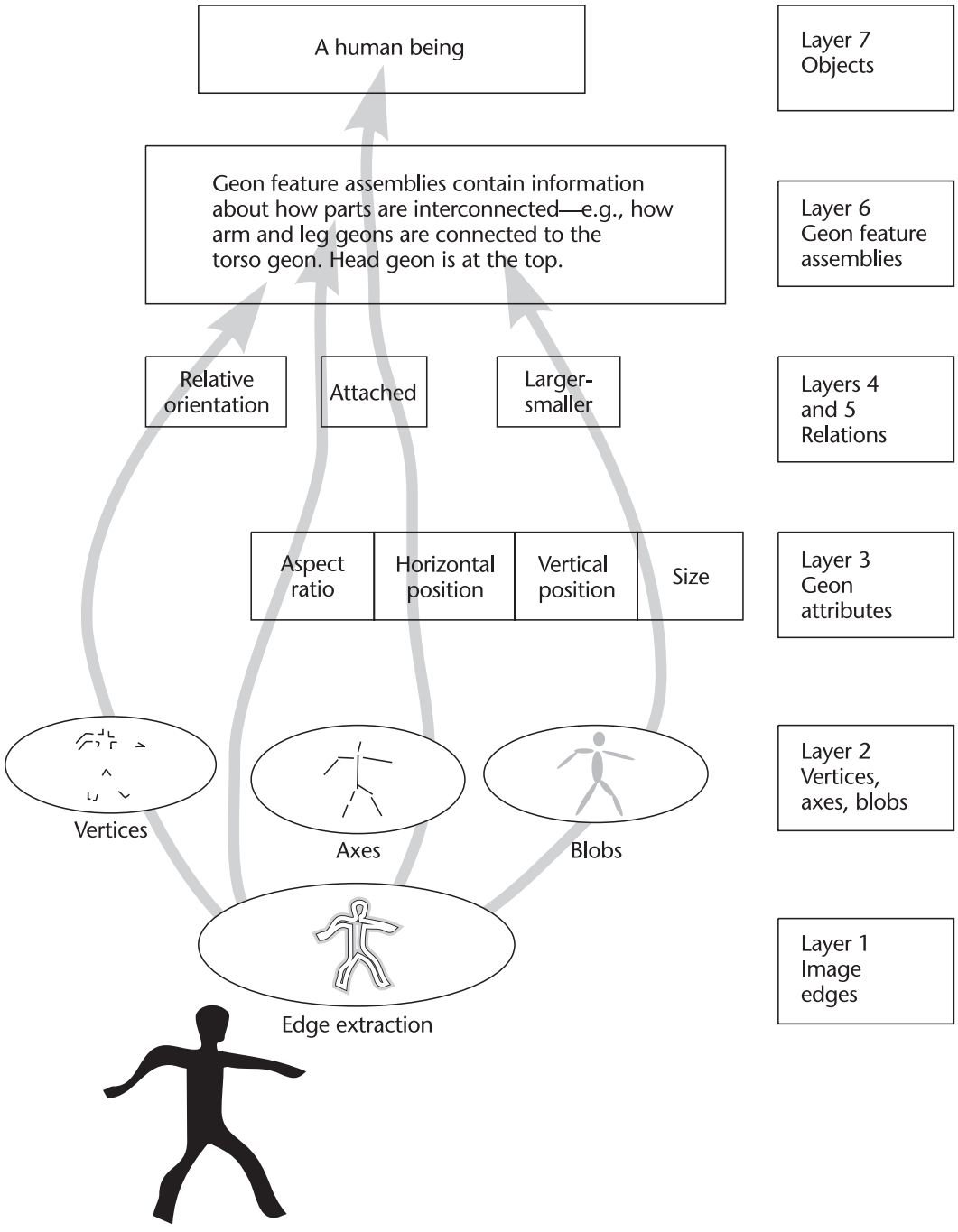

Figure 7.4 A simplified view of Hummel and Biederman's (1992) neural-network model of form perception.

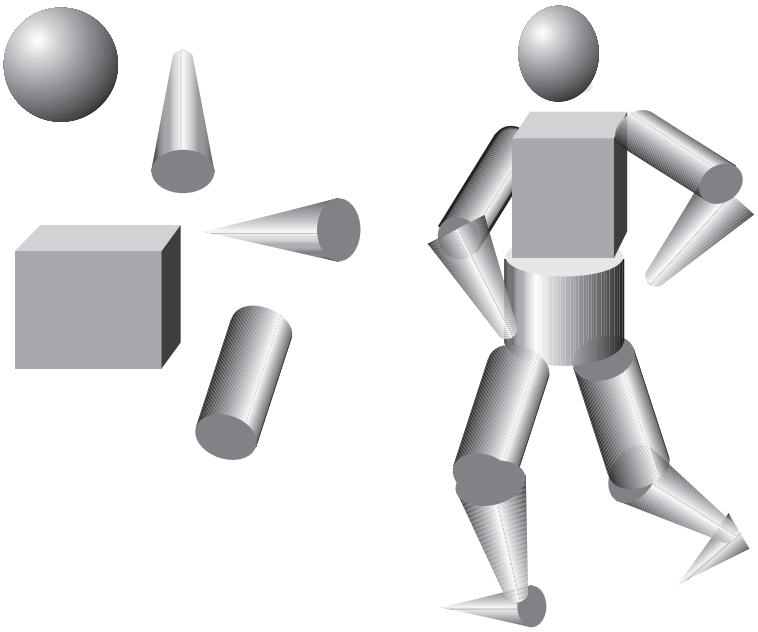

Figure 7.5 According to Biederman's geon theory, the visual system interprets 3D objects by identifying 3D component parts called geons.

silhouettes that are easily recognizable; think of a teapot, a shoe, a church, a person, or a violin. These *canonical* silhouettes are based on a particular view of an object, often from a point at right angles to a major plane of symmetry. Figure 7.6 illustrates canonical views of a teapot and a person.

David Marr suggested ways in which the brain might use silhouette information to extract the structures of objects (Marr, 1982). He argued that "buried deep in our perceptual machinery" are mechanisms that contain constraints determining how silhouette information is interpreted. Three rules are embedded in this perceptual machinery:

1. Each line of sight making up a silhouette grazes the surface exactly once. The set of such points is the *contour generator*. The idea of the contour generator is illustrated in Figure 7.7.
2. Nearby points on the contour of an image arise from nearby points on the contour generator of the viewed object.
3. All the points on the contour generator lie on a single plane.

Under Marr's default assumptions, contour information is used in segmenting an image into its component solids. Marr and Nishihara (1978) suggested that concave sections of the silhouette contour are critical in defining the ways different solid parts are perceptually defined. Figure 7.8 illustrates a crudely drawn animal that we nevertheless readily segment into head, body, neck,

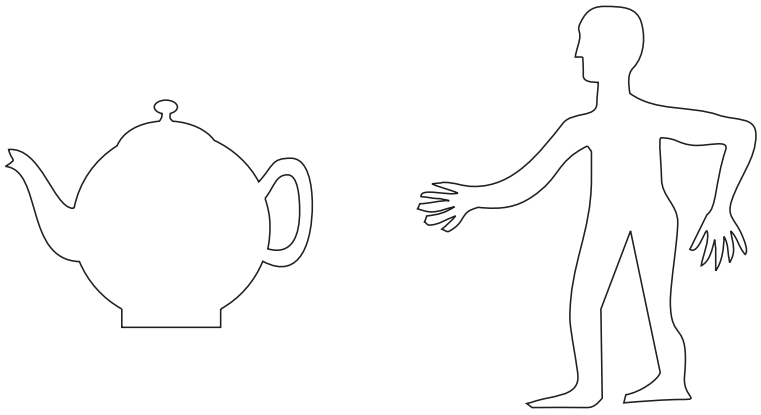

Figure 7.6 Many objects have canonical silhouettes, defined by the viewpoints from which they are most easily recognized. In the case of the man, the overall posture is unnatural, but the component parts—hands, feet, head, and so on—are all given in canonical views.

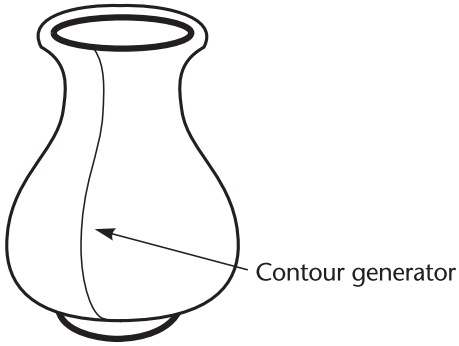

Figure 7.7 According to Marr, the perceptual system makes assumptions that occluding contours are smoothly connected and lie in the same plane. *Adapted from Marr (1982).*

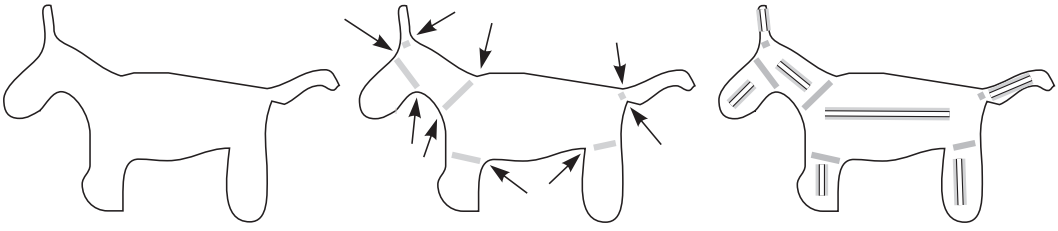

Figure 7.8 Concave sections of the silhouette define subparts of the object and are used in the construction of a structural skeleton. *Adapted from Marr and Nishihara (1978).*

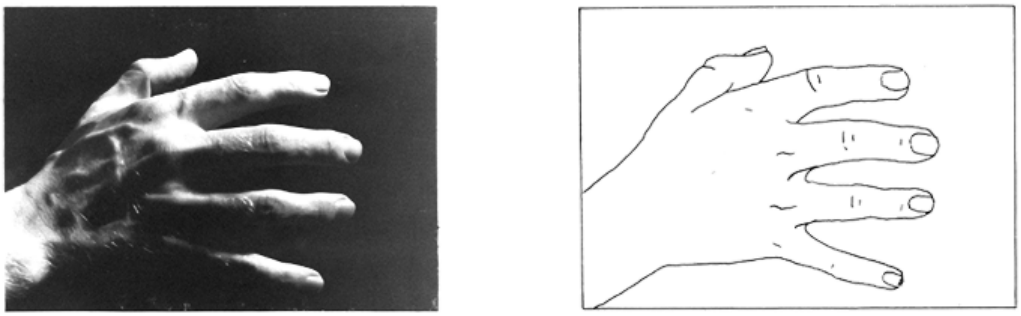

Figure 7.9 A photograph of a hand and a simplified line drawing of the hand. Ryan and Schwartz (1956) showed that a cartoon image was recognized more rapidly than a photograph.

legs, and so on. Marr and Nishihara also suggested a mechanism whereby the axes of the parts become cognitively connected to form a structural skeleton.

One of the consequences of structural theories of perception is that certain simplified views should be easier to read. There are practical advantages to this. For example, a clear diagram may sometimes be more effective than a photograph. This is exactly what Ryan and Schwartz (1956) showed when they found that a hand could be perceived more rapidly in the form of a simplified line drawing than in the form of a photograph (see Figure 7.9).

But this result should not be overgeneralized. Other studies have shown that time is required for detailed information to be perceived (Price and Humphreys, 1989; Venturino and Gagnon, 1992). Simplified line drawings may be most appropriate only when rapid responses are required.

Although image-based theories and structure-based theories of object recognition are usually presented as alternatives, it may be that both kinds of processes occur. If geons are extracted based on concavities in the silhouette, certain views of a complex object will be much easier to recognize. Further, it may well be that viewpoint-dependent aspects of the visual image are stored in addition to the 3D structure of the object. Indeed, it seems likely that the brain is capable of storing many kinds of information about an object or scene if they have some usefulness. The implication is that even though 3D objects in a diagram may be more effective in some cases, care should be taken to provide a good 2D layout.

Faces

Faces are special objects in human perception. Infants learn about faces faster than other objects. It is as if we are born with visual systems primed to learn to recognize important humans, such as our own mothers (Morton and Johnson, 1991; Bruce and Young, 1998; Bushnell et al., 1989). A specific area of our brains, the right middle fusiform gyrus, is especially important in face perception (Puce et al., 1995; Kanwisher et al., 1999; Kanwisher et al., 1997). This area is

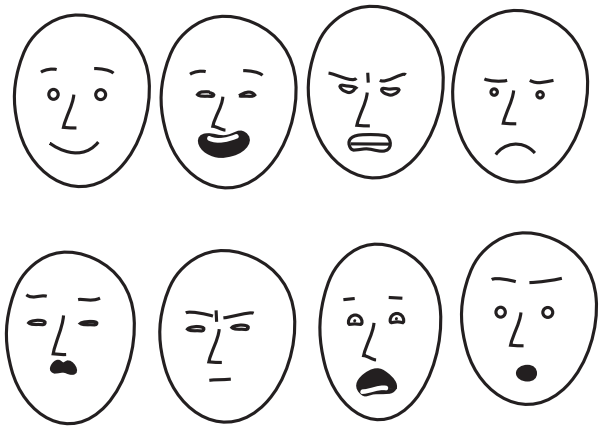

Figure 7.10 Happiness, elation, anger, sadness, disgust, determination, fear, surprise.

also useful for recognizing other complex objects, such as automobiles; although it is not essential as a Volkswagen detector, we cannot recognize faces without it.

Faces have an obvious importance in communication, because we use facial expression to communicate our emotion and degree of interest. Cross-cultural studies by Paul Ekman and coworkers strongly suggests that certain human expressions are universal communication signals, correctly interpreted across cultures and social groups (Ekman and Friesen, 1975; Ekman, 2003). Ekman identified six universal expressions: anger, disgust, fear, happiness, sadness and surprise. These are illustrated in Figure 7.10, along with determination and elation (a variation on happiness). The motion of facial features is also important in conveying emotion. Animated images are necessary to convey a full range of nuanced emotion; it is especially important to show motion of the eyebrows (Basilli, 1978; Sadr et al., 2003).

Facial expressions are produced by the contractions of facial muscles. The *facial action coding system* (FACS) is a widely applied method of measuring and defining groups of facial muscles and their effect on facial expression (Ekman et al., 1988). The eyebrows and mouth are particularly significant in emotion signaling, but the shape of the eyes is also important. There is evidence that false smiles can be distinguished from true smiles from the particular expression around the eyes that occurs with contraction of a muscle that orbits the eye (Ekman et al., 1988; Ekman, 2003). This muscle contracts with true smiles but not with false ones. According to Ekman (2003) it is difficult, if not impossible, to control this voluntarily and thus fake a "true" smile.

The main application of FACS theory in computer displays has been in the creation of computer avatars that convey human emotion (Kalra et al., 1993; Ruttkay et al., 2003). Appropriate emotional expression may help make a virtual salesperson more convincing. In computer-aided instruction, the expression on a human face could reward or discourage.

The Object Display and Object-Based Diagrams

Wickens (1992) is primarily responsible for the concept of an *object display* as a graphical device employing "a single contoured object" to integrate a large number of separate variables. Wickens theorized that mapping many data variables onto a single object will guarantee that these variables are processed together, in parallel. This approach, he claimed, has two distinct advantages. The first is that the display can reduce visual clutter by integrating the variables into a single visual object. The second is that the object display makes it easier for an operator to integrate multiple sources of information.

Among the earlier examples of object displays are Chernoff faces, named after their inventor, Herman Chernoff (1973). In this technique, a simplified image of a human face is used as a display. Examples are shown in Figure 7.11. To turn a face into a display, data variables are mapped to different facial features, such as the length of the nose, the curvature of the mouth, the size of the eye, the shape of the head, etc. There are good psychological reasons for choosing what might seem to be a rather whimsical display object. Faces are probably the most important class of objects in the human environment. Even newborn babies can rapidly distinguish

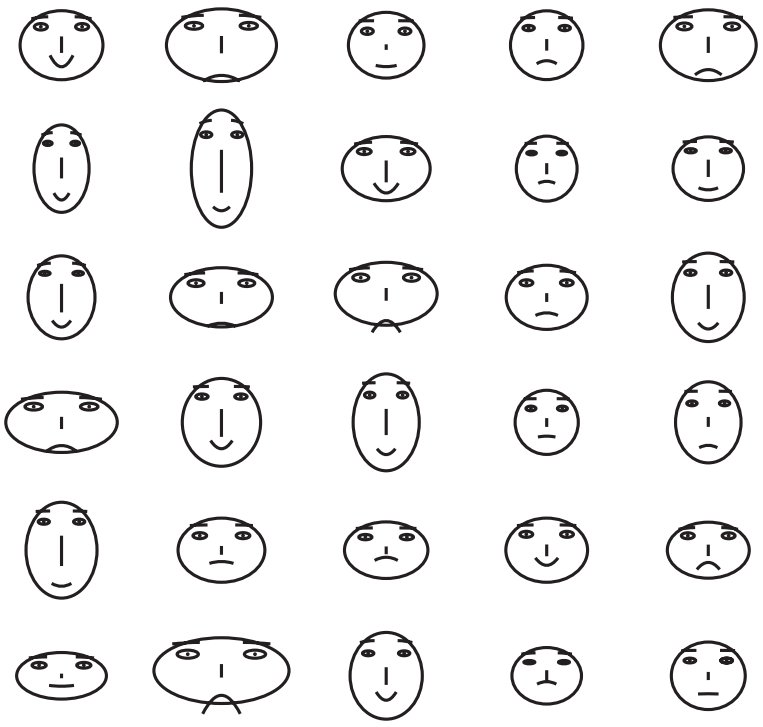

Figure 7.11 Chernoff faces. Different data variables are mapped to the sizes and shapes of different facial features.

faces from nonfaces with scrambled features, suggesting that we may have special neural hardware for dealing with faces. Jacob et al. (1976) carried out a classification task using a series of displays that were progressively more objectlike. The displays included Chernoff faces, tables, star plots, and the whisker plots described in Chapter 5. They found that the more objectlike displays, including Chernoff face plots, enabled faster, more accurate classification.

Chernoff faces have not generally been adopted in practical visualization applications. The main reason for this may be the idiosyncratic nature of faces. When data is mapped to faces, many kinds of perceptual interactions can occur. Sometimes the combination of variables will result in a particular stereotypical face, perhaps a happy face or a sad face, and this will be identified more readily. In addition, there are undoubtedly great differences in our sensitivity to the different features. We may be more sensitive to the curvature of the mouth than to the height of the eyebrows, for example. This means that the perceptual space of Chernoff faces is likely to be extremely nonlinear. In addition, there are almost certainly many uncharted interactions between facial features, and these are likely to vary from one viewer to another.

Often, object displays will be most effective when the components of the objects have a natural or metaphorical relationship to the data being represented. For example, Figure 7.12 illustrates how a storage vessel in a chemical plant might be represented using both a conventional bar chart and a customized object display. The variables in the object diagram are represented as follows:

- Size of cylinder represents tank capacity.
- Height of liquid represents volume of material stored.
- Texture of liquid represents the chemical composition.
- Color of liquid represents liquid temperature.
- Diameter of pipe represents outflow capacity.
- Status of the valve and thickness of the outgoing fluid stream represent rate at which liquid is being drawn from the tank.

In this example, the object display has a number of clear advantages. It can reduce accidental misreadings of data values. Mistakes are less likely because components act as their own descriptive icons. In addition, the structural architecture of the system and the connections between system components are always visible, and this may help in diagnosing the causes and effects of problems. Conversely, the disadvantage of object displays is that they lack generality. Each display must be custom-designed for the particular application and, ideally, should be validated with a user population to ensure that the data representation is clear and properly interpreted. This requires far more effort than displaying data as a table of numbers or a simple bar chart.

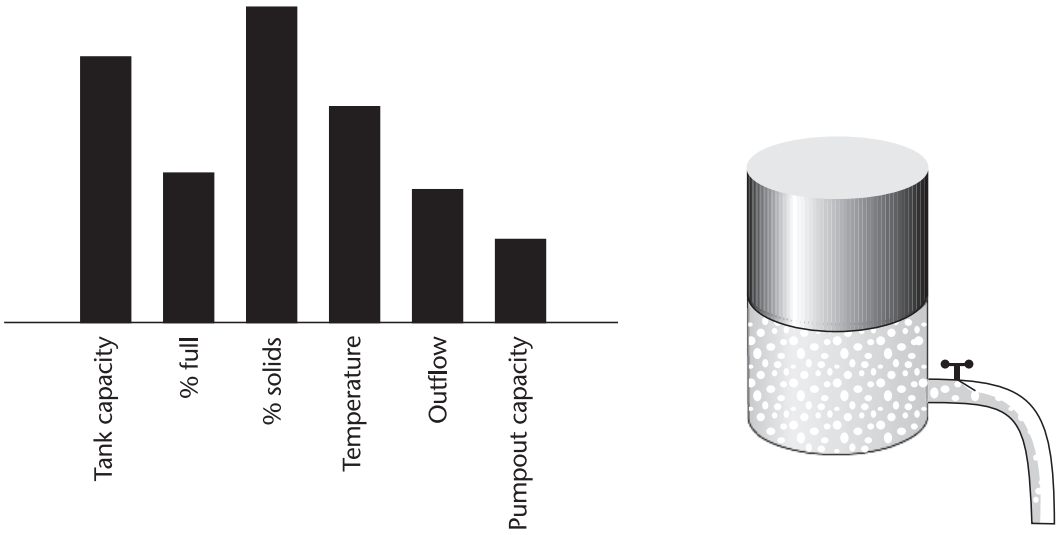

Figure 7.12 Two representations of the same data. The object diagram on the right combines six variables in an easily interpreted, cohesive representation.

The Geon Diagram

Biederman's geon theory, outlined earlier, can be applied directly to object display design. If cylinders and cones are indeed perceptual primitives, it will make sense to construct diagrams using these geon elements. This should make the diagrams easy to interpret if a good mapping can be found from the data to a geon structure. The geon diagram concept is illustrated in Figure 7.13(a). Geons are used to represent the major components of a compound data object, whereas the architecture of the data object is represented by the structural skeleton linking the geons. The size of a geon becomes a natural metaphor for the relative importance of a data entity, or its complexity or relative value. The strength of the connections between the components is given by the necklike linking structures. Additional attributes of entities and relationships can be coded by coloring and texturing them.

We evaluated the geon diagram concept in a comparison with Unified Modeling Language (UML) diagrams (Irani et al., 2001). UML is a widely used, standardized diagramming notation for representing complex systems. Equivalent diagrams were constructed by matching geon elements to UML elements (see Figure 7.13). We found that when the task involved rapid identification of substructures in a larger diagram, participants performed both faster and with only half the errors using the geon diagrams. Another experiment showed that geon diagrams were easier to remember.

In Biederman's theory, surface properties of geons, such as their colors and textures, are secondary characteristics. This makes it natural to use the surface color and texture of the geon to represent data attributes of a data object. The important mappings between data and a geon diagram are as follows:

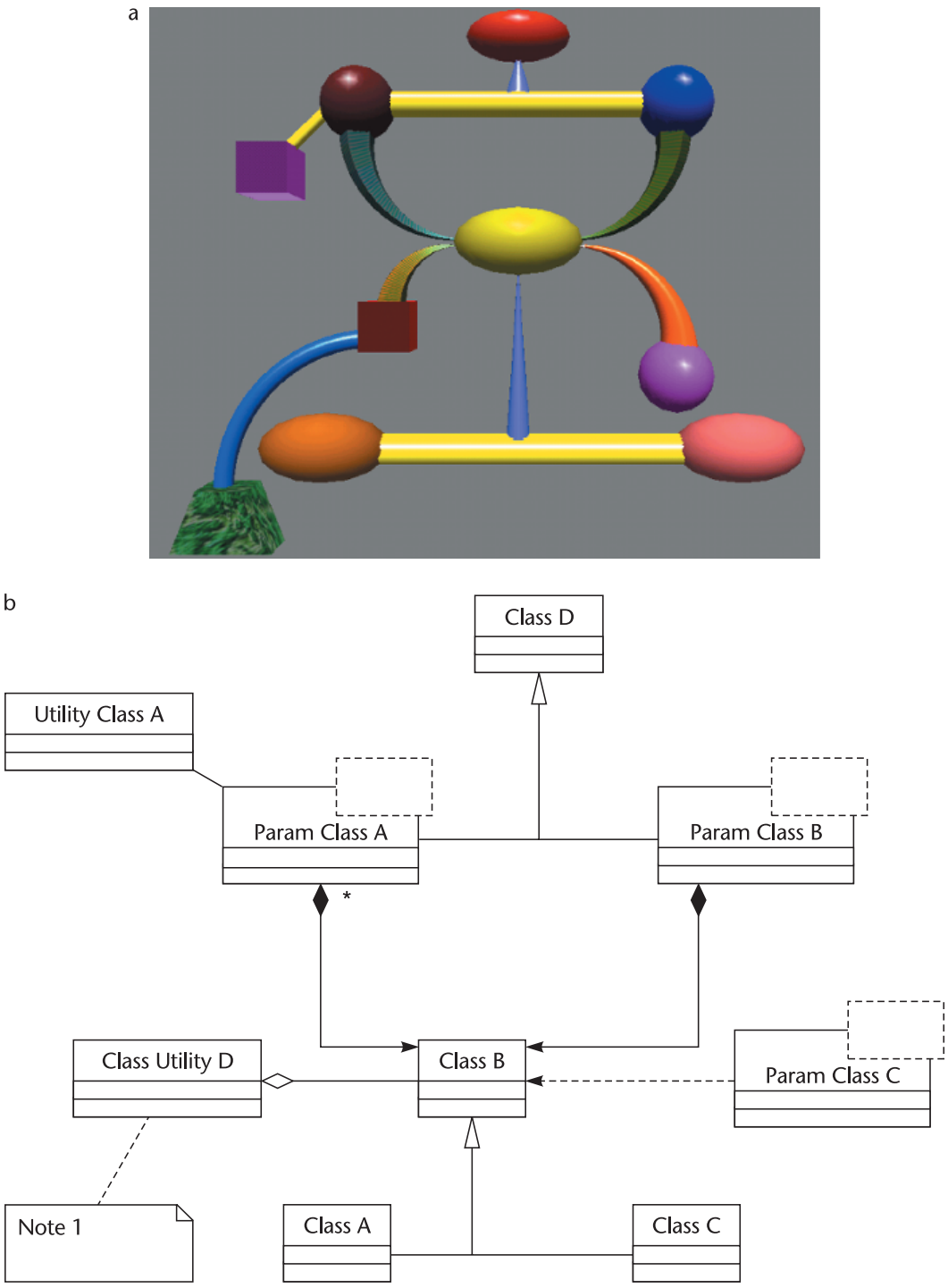

Figure 7.13 (a) A geon diagram constructed using a subset of Biederman's geon primitives. The primitive elements can also be color-coded and textured. (b) A Unified Modeling Language (UML) equivalent.

Major components of a complex data object	→	Geons
Architectural links between data object components	→	Limbs consisting of elongated geons—connections between limbs reflect architectural structure of data
Minor subcomponents	→	Geon appendices—small geon components attached to larger geons
Component attributes	→	Geon color, texture, and symbology mapped onto geons

Although the geon diagram is a 3D representation, there are reasons to pay special attention to the way it is laid out in 2D in the *x,y* plane. As discussed earlier, some silhouettes are especially effective in allowing the visual system to extract object structure. Thus, a common-sense design rule is to lay out structural components principally on a single plane. A diagramming method resembling the bas-relief stone carvings common in classical Rome and Greece may be optimal. Such carvings contain careful 3D modeling of the component objects, combined with only limited depth and a mainly planar layout.

Abstract semantics may be expressible, in a natural way, through the way geons are interconnected. In the everyday environment, there is meaning to the relative positioning of objects that is understood at a deep, possibly innate level. Because of gravity, *above* is different from *below*. If one object is inside another, it is perceived as either contained by that other object or a part of it. Irani et al. (2001) suggested that the semantics inherent in the different kinds of relationships of real-world objects might be applied to diagramming abstract concepts. Based on this idea, the researchers developed a set of graphical representations of abstract concepts. Some of the more successful of these mappings are illustrated in Figure 7.14 and listed as follows.

- Sometimes we wish to show different *instances* of the same generic object. Geon theory predicts that having the same shape should be the best way of doing this. Geon shape is dominant over color, which is a secondary attribute. Thus the elbow shapes in Figure 7.14(a) are seen as two instances of the same object, whereas the two green objects are not.
- Having an object inside another transparent object is a natural representation of a *part-of* relationship. The inside objects seem part of the outside objects, as in Figure 7.14(b).
- One object above and touching another, as shown in Figure 7.14(c), is easily understood as representing a *dependency* relationship.
- A thick bar between two objects is a natural representation of a *strong* relationship between two objects; a thinner, transparent bar represents a *weak* relationship. See Figure 7.14(d).

Perceiving the Surface Shapes of Objects

Not all things in the world are made up of closed, discrete components like geons. For example, there are undulating terrains that have no clearly separable components. Although to some extent