From Kosslyn et al. 2006 *The Case for Mental Imagery*

**Table 1.1**
(*Continued*)

| | Propositional | Depictive |
|---|---|---|
| | Abstract:<br>Can refer to non-picturable entities | Cannot be abstract in any sense:<br>Refers to picturable entities |
| | Can refer to classes of objects | Refers to exemplars |
| | Not tied to any specific sensory modality | Specific to a particular sensory modality |
| Rules of Combination | | |
| | Arrangement of symbols can assert a true or false state of affairs | As spatial relations among points are distorted, the depiction becomes increasingly less accurate |

the right leg, and so on. Such spatial relations are immediately accessible in a depictive representation, and thus depictive representations may be more efficient in representing the spatial layout of objects. On the other hand, the fact that there are three "lines" (i.e., strokes) is explicit in the descriptive representation but not in the depictive one. In the depiction, all portions of the pattern are both explicit and accessible, but they are not semantically interpreted; thus, to access the concept of "line," a process must interpret the pattern. In contrast, the propositional representation makes this interpretation of the pattern explicit and accessible.

In sum, depictive representations make explicit and accessible all aspects of shape and the relations between shape and other perceptual qualities (such as color and texture), as well as the spatial relations among each point. In contrast, propositional representations make explicit and accessible semantic interpretations, which can include aspects of shape and other perceptual qualities. Depictive representations of shape must also incidentally specify size and orientation; propositional representations only specify what was explicitly included when the representation was created. Depending on the precise task at hand, one or the other format may be most useful.

## Depictions in the Brain

As we saw in the computer analogy described above, there need be no actual picture in the brain to have a depiction: all that is needed is a "functional space" in which distance can be defined vis-à-vis how information is processed. (However, as we discuss in chapter 2, a functional space is sufficient as a depictive form of representation only if the geometric properties of the representation emerge because there are fixed, hard-wired processes that interpret the representation as if it were a space; if the processes are not fixed, then the representation is not necessarily a depiction.)

Nevertheless, even though all that is required in order to have a depiction is a functional space, there is good evidence that the brain depicts representations literally, using space on the cortex to represent space in the world. To be specific, we will argue in the following chapters that images rely in part on areas in the brain that are specifically designed to depict patterns. These areas are *topographically organized*—they preserve (roughly) the geometric structure of the retina. Such areas use space on the cortex to represent space in the world (e.g., see Felleman & Van Essen, 1991; Fox et al., 1986; Heeger, 1999; Sengpiel & Huebener, 1999; Sereno et al., 1995; Tootell, Hadjikani, Mendola, Marrett, & Dale, 1998; Tootell, Silverman, Switkes, & De Valois, 1982; Van Essen, 1985). For example, figure 1.2 illustrates the results of an experiment reported by Tootell, Silverman, Switkes, and De Valois in 1982. They trained a monkey to stare at the pattern shown on the left, which consisted of a set of blinking lights arranged as shown. The animal was injected with a radioactive form of sugar, which was taken up into brain cells in proportion to how active the cells were while the animal observed the pattern; the more active the brain cell, the more sugar it took up. This particular isotope gets lodged in the neurons and is not quickly broken down by metabolic processes. The animal was then sacrificed, and its brain was removed. Figure 1.2 illustrates the first cortical area to receive input from the eyes, known variously (but synonymously) as area V1, area 17, area OC, the striate cortex, and the primary visual cortex. The dark bands in the right part of figure 1.2 label brain cells that took up a lot of the radioactively tagged sugar. As is clear, the geometric structure of the stimulus is physically laid out on the cortex!

These areas do not simply have a topographically organized physical structure; they *function* to depict information. If a patch of cortex in one of these areas is damaged (for example, because a tumor had to be removed), this damage will produce a scotoma (i.e., a "blind spot") in the corresponding part of the visual field. The scotomas that arise when the topographically organized visual cortex is damaged demonstrate conclusively that these areas function to depict information; crucially, the closer two damaged regions of the topographically organized visual cortex are, the closer in the visual field the corresponding scotomas will be. And this result is not simply about the effects of chronic damage: transcranial magnetic

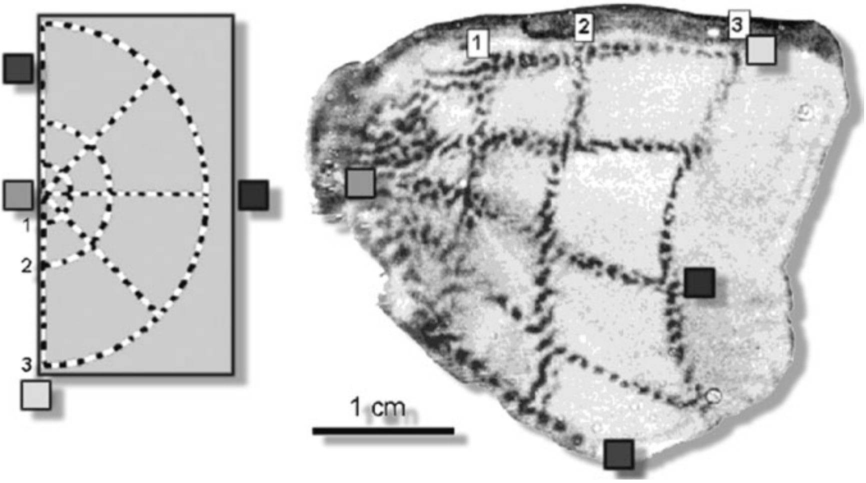

**Figure 1.2.** A geometric pattern of flashing lights (*left*) shown to a monkey after it was injected with a radioactive sugar, which is taken up into the brain cells in proportion to their level of activity. The animal was trained to stare at the stimulus, and then was sacrificed so that its brain could be examined. The occipital cortex of the monkey (*right*) is shown with dark bands revealing the neurons that were most activated while the animal viewed the pattern. A pattern of activation corresponding to the geometric structure of the stimulus can be seen clearly. This result demonstrates that visual stimuli are represented topographically in the occipital cortex of the monkey brain. Shaded squares are used to demarcate the points on the visual stimulus that are represented in corresponding space on the visual cortex. The numbers 1, 2, and 3 indicate the inner, middle, and outer bands of flashing lights that are clearly represented on the cortex. Note that for simplicity, only half the circular light stimulus is shown here, along with a section of V1 from a single hemisphere of the brain. (Adapted with permission from Tootell et al. [1982]. *Science, 218*, 902–904; and Tootell et al. [1988]. *Journal of Neuroscience, 8*[5], 1531–1568.)

stimulation has been used to stimulate occipital cortical sites transiently to produce phosphenes (i.e., bright flashes of light that are not produced by sensory input); when nearby sites are stimulated, phosphenes appear in nearby locations in space, and when far apart sites are stimulated, phosphenes appear in far apart locations in space (e.g., Kastner, Demmer, & Ziemann, 1998).

The appearance of figure 1.2 notwithstanding, the pattern of activation is not literally a picture of what the animal is seeing. From the point of view of an outside observer (such as the reader, observing that illustration), metric information on the cortex is distorted. However, as we discuss at length in the following pages, the key is how processes that receive outputs from this cortex interpret distance on the cortex—not how a human observer armed with a ruler would interpret such distances. If we were to treat the cortex like a piece of paper, we would conclude that the depiction is distorted, with the central part amplified—but the brain processes that access this representation correct for such distortions. Within the context of the brain as a whole, these topographically organized areas truly depict information.

Why does the brain use space on the cortex to represent space in the world? Although the ease of genetic coding or other factors may play a role, the best current guess is that this structure has been retained through evolution for a simple reason: this trick makes explicit and accessible information needed for the tasks at hand.[5] For example, the first cortical visual areas to receive input from the eyes are confronted with the task of organizing figure from ground. In order to do so, they must delineate edges. This task is facilitated by the fact that many (perhaps most) of the connections among neurons in a topographically organized area are both very short and inhibitory. This means that if one neuron is stimulated, it attempts to inhibit its neighbors—which represent adjacent points in space. The effect of this is to exaggerate differences in neural activation across the boundaries of edges. For instance, if an object reflects a lot of light and the background does not, then neurons that are near the edges of objects will inhibit those that register the background, which will amplify the difference in their activation—thereby helping to identify edges of objects.

For mental imagery, purely functional depictions may have been sufficient; there is no obvious reason why physical depictions are required. However, because the imagery system draws on mechanisms used in like-modality perception (as we demonstrate throughout the remainder of this book), it relies on such physically topographically organized structures. As we argue, many of the properties of imagery arise because of this simple fact. For instance, because the input to early topographically organized areas changes every time the eyes move, the patterns of activation within them cannot linger for long; if they did, the world would seem smeared as we moved our eyes. But what is a virtue for perception is a drawback for imagery: as we argue in chapter 5, it is difficult to maintain images for long, in part because they rely on neural machinery also used in perception.

In short, each location on the topographically organized cortex corresponds to a specific location in space, and distance between the locations on cortex corresponds to distance between the corresponding locations in space. The brain has numerous such representations, but we shall emphasize two classes of them in this book. First, many of these depictive representations are involved in processing shapes, particularly in the occipital lobe. We group the topographically organized areas in the occipital lobe into a single functional structure, which we refer to as the *visual buffer*. Patterns of activation within the visual buffer depict shapes, according to the definition of *depiction* offered here. The visual buffer, in essence, is the canvas upon which images are painted; it is the medium that supports depictive representations. In later chapters we provide strong evidence that topographically organized brain areas are in fact used in visual mental imagery, but first we will use the idea that imagery relies on patterns of activation within such areas to show that the arguments levied against depictive representations do not hold water.

Second, the brain also uses depictive representations to specify information about the locations of objects in space. These representations are primarily in the posterior parietal lobes (e.g., Sereno, Pitzalis, & Martinez, 2001). Although a depictive format is used both to represent shape and location, we shall see that the contents of these representations differ markedly. As we discuss in later chapters, because of the differences in content, we will distinguish between *object images,* which represent shape (and shape-related properties, such as color and texture), versus *spatial images*, which represent relative locations in space.

In the most recent round of the debate, at stake is the very idea that turning to the brain can inform theories of cognition. We will argue that the function of the brain—which is, after all, the organ of thought—has evolved in tandem with its structure, and vice versa: the brain is not a "general purpose computer," like a von Neumann machine (such as your personal computer), which acts very differently depending on the program it has in memory. Instead, the brain is largely a special-purpose machine, which is tailored to function in specific ways. Because different formats make different information explicit and accessible, different formats are more or less useful for performing different tasks. Thus, an efficient strategy for the brain is to use different formats in different situations, and animals that could do this may have had an adaptive advantage over their less specialized brethren.

## Hybrid Depictive Representations

Consider again the gripping illustration in figure 1.2. Those dark lines indicate where neurons were particularly active while the monkey was observing the pattern. If we take this finding to indicate that a depictive representation is being used (as we should, given our characterization of such representation), we are led to conclude that mental images are not like

# 4

# Depictive Representations in the Brain

The imagery debate hinges on the distinction between the format and the content of internal representations, the format being the type of code used to represent the shape of an object, and the content being the information conveyed. The issue is whether mental imagery makes use of representations that depict (perhaps in addition to those that also describe). Following many years of behavioral research, this debate reached an impasse: the results from these studies could be explained by theories that posited only propositional representations or by theories that also posited depictive representations (e.g., see Anderson, 1978). Modern neuroimaging techniques now provide an opportunity to press this issue forward, for two reasons.

First, researchers have documented that approximately thirty-two areas of the monkey cerebral cortex contain neurons that respond selectively to visual input; moreover, about half of these areas are topographically organized—they preserve (roughly) the geometric structure of the retina. Such areas use space on the cortex to represent space in the world (e.g., see Felleman & Van Essen, 1991; Fox et al., 1986; Heeger, 1999; Sengpiel & Huebener, 1999; Sereno et al., 2001; Tootell, Hadjikani, et al., 1998; Tootell, Silverman, et al., 1982; Van Essen, 1985). As discussed in chapter 1, these areas are not simply physically topographically organized—they function to depict information. The representations in topographically organized areas are depictions, not propositional descriptions.

Positron emission tomography (PET; e.g., Fox et al., 1986) and functional magnetic resonance imaging (fMRI; e.g., Sereno et al., 1995) have shown that such depictive areas are used during visual perception in the human brain. If these areas are activated when one visualizes, and disruption of these areas impairs the ability to visualize, this is strong evidence that the representations underlying visual mental images are not entirely propositional; rather, such evidence would show that at least some of these representations depict information and that such representations play a role in information processing.

The second reason neuroimaging results offer a way to break the impasse in the debate about imagery is that such findings cannot be explained by appeal to tacit knowledge. As we discussed in the previous chapter, some researchers have claimed that the results from studies of imagery arise because task demands are built into the experiments, which in turn activate tacit knowledge about perceived events (a phenomenon that is largely unconscious, such as our knowledge of how the surface of water will appear when a glass of water is tilted), which in turn leads participants to produce responses like those they believe (perhaps unconsciously) would occur in the corresponding perceptual situation. Findings from the brain would be truly definitive evidence against this position: no ordinary person has knowledge of how perception (and hence, according to the tacit knowledge account, also imagery) is processed in the brain—and thus people are not in a position to manipulate their brain activation (even if they had that ability) in accordance with tacit knowledge.

In the first part of this chapter, we summarize the arguments against the idea that topographically organized areas are used in visual mental imagery and the arguments that even if they are used, they play no essential role. In the second part, we address each argument in turn, considering counterarguments and evidence to the contrary. In the third part, we discuss a meta-analysis of studies of visual mental imagery (summaries of the studies are provided in the appendix). This meta-analysis untangles what might at first blush appear to be inconsistencies in the literature and provides strong support for the claim that topographically organized areas support depictive representations during visual mental imagery.

## Do Topographically Organized Areas Depict Information?

Some researchers have tried to dismiss the evidence that depictive representations are functional in the brain. Such arguments can be divided into three classes, summarized below. We first summarize the arguments, and then consider them in detail.

### Arguments against Neural Depiction

Some critics of the recent neuroimaging research have claimed that topographically organized areas in the human brain are not truly depictive. For example, area 17 (also known as area V1, the striate cortex, and the primary visual cortex) is organized so that objects that subtend small visual angles activate posterior portions of the area, and objects that subtend larger visual angles activate more anterior portions of the area (e.g., Fox et al., 1986). Larger objects do not necessarily occupy a larger amount of the area, so how can we infer that the area is truly depicting information?

In addition, the idea that topographically organized regions in the brain support depictive representations purportedly is flawed because vision does not accumulate information in an internal display to produce a panoramic view of the world; this is a subjective illusion. Every time the eyes move, the information in early visual structures is replaced by the new input. So, how can theorists claim that neural structures used in vision also support depictive representations in mental imagery? In criticizing our claim that topographically organized visual areas depict information, Dennett (2002, p. 190) asserts: “It has yet to be established when and how *vision* utilizes image processing!” He continues:

> Vision isn’t television. The product of vision is not a picture on the screen in the Cartesian Theater (Dennett 1991). The fleeting retinal images punctuated by saccades are the first images, and they are not the last, as Julesz (1971) demonstrated by showing perception of depth in random dot stereograms that requires image-processing after the optic chiasma. But which subsequent cortical processes also exploit any of the informational properties of images? The *eventual* “products” of vision are such things as guided hand and finger motions, involuntary ducking, exclamations of surprise, triggering of ancient memories, sexual arousal . . . and none of these is imagistic in any sense, so assuming that the events in their proximal causal ancestry are imagistic is rather like assuming that power from a hydroelectric plant is apt to be wetter and less radioactive than power from a nuclear plant. The raw retinal data are cooked in many ways betwixt eyeball and verbal report (for instance). How cooked are the processes involved

> in (deliberate or voluntary) mental imagery? (Dennett, 2002, p. 190)

Moreover, continuing this line of argument, critics have claimed that the very idea that the brain would create a copy of the visual world is flawed. This idea is nicely captured in the following quotation: "There is no evidence of the mental construction of images to be looked at or maps to be followed. The body responds to the world, at the point of contact; making copies would be a waste of time." These words were written by B. F. Skinner (1977, p. 6) but could have been written by many other critics of depictive theories of imagery. Pylyshyn (2002, p. 182, n. 14) asserts that vision "is not for turning one topographically organized pattern into another." Echoing this point, Dennett (2002) emphasizes that vision is used to identify objects and their properties and to guide and track movements—and none of these eventual goals directly relies on topographic maps. Dennett (2002) notes that electricity generated from hydroelectric plants is not wetter than electricity generated by other means—and thus, by analogy, topographically mapped areas may have no bearing on the representations actually used in vision.

## Clarifying the Nature of Depictive Representation

Here we discuss why the arguments raised against depictive representations in the brain miss their intended target.

### Representation within Topographically Organized Areas

Many neuroimaging findings have documented topographic organization in several areas of human visual cortex during perception (e.g., DeYoe, Bandettini, Neitz, Miller, & Winans, 1994; Engel et al., 1994; Hasnain, Fox, & Woldorff, 1998; Sereno et al., 1995; Van Essen et al., 2001). The high-resolution central (foveal) portions of these areas represent the central portions of the visual field, whereas low-resolution, peripheral portions of these areas represent more peripheral portions of the visual field. Space on the cortex is literally used to represent space in the world (more precisely, space on the cortex is used to represent a planar projection of space in the world, because the cortical areas are two-dimensional; information about depth is represented in a different manner, e.g., Tsao et al., 2003). Propositional representations have the property of being arbitrarily related to the objects they represent, whereas the relationship between an external visual stimulus and the activation it engenders in topographically organized areas is not arbitrary: points that are relatively close in space in the external world are represented by activation in nearby portions of these areas, and points that are relatively far in space are represented by activation in relatively separated portions within each of these areas.[1]

What about the question of why larger objects do not produce larger swaths of activation within these brain areas? This concern is based on a misunderstanding of how information is represented in topographically organized areas. First, neurons in the anterior portions of area 17 have larger receptive fields, which receive inputs from a larger region of space than do neurons in more posterior portions. Thus, activation in more anterior portions of this structure indicates that parts of the image are spaced farther apart and are less resolved than is the case when activation is in more posterior portions. This point is worth repeating, since it is often not appreciated: activation toward more anterior portions of this structure not only depicts larger extents but does so with poorer resolution. However, keep in mind that what is important is not how the area looks to an external observer but how it is interpreted by processes that operate on it in the brain. From this perspective, when neurons in the anterior portions of area 17 are activated, this is interpreted as specifying a larger portion of an object's surface.

The spatial organization of topographically organized areas in the human brain is complex, in part because numerous topographically organized areas must be connected to preserve the overall topography. Adjacent topographically organized areas meet at the horizontal meridian (i.e., the horizontal line through the fixation point in the frontal plane of one's visual field) or at the vertical meridian (i.e., the vertical line through the fixation point in the frontal plane). The most common way to establish that areas are topographically organized takes advantage of these facts. In this method, participants observe a narrow flashing wedge as it rotates slowly around a pivot, starting from the vertical. fMRI is used to track changes in activation over time as the stimulus moves. If the stimulus is vertical at the outset, activation will initially be detected along the vertical meridian, and as the stimulus rotates, this activation literally shifts over the cortex, toward the horizontal meridian. Because adjacent visual cortical areas abut at the horizontal (or vertical) meridian, this means that the wave of activation in one area will literally shift until it meets an incoming wave of activation from an abutting area—and researchers use the places where waves meet to establish the boundaries between topographically organized areas. We wish to emphasize that these waves reflect the precise position of the stimulus in space: as the stimulus moves, the activation is mapped, point-for-point, into the topographically organized areas. Each neuron in these areas registers a part of space, and the distance between neurons reflects the distance between the corresponding parts of space. However, the physical distance on cortex also reflects the resolution—with the same extent on an object's being represented by greater distances in the foveal region of cortex than in peripheral regions.[2]

And this brings us to the crucial point: connections to other areas compensate for the distortions in the actual (physical) representation—and thus the result is that the topographic areas function to depict shape accurately. This is a variant of the "Modigliani effect," named after the Italian artist Amedeo Clemente Modigliani (1884–1920), who famously painted and sculpted elongated figures. The question was once raised whether the painter actually saw the world the way he portrayed it. However, if Modigliani had in fact had distorted vision, which led him to see objects as elongated, the same distortions would operate when he painted, thereby compensating and leading him to produce a veridical depiction. But in the case of the brain, later brain areas compensate for the distortions introduced into earlier visual areas. Topographically organized areas depict, but do so within the context of the system as a whole.

## The Role of Topographically Organized Areas in Perception

We must distinguish two separate issues. First, are depictive representations actually used in vision at all, no matter whether it is for a single fixation or only a portion of the visual field? Second, if so, how are such representations combined successively to produce the impression of seeing the entire visual field?

Regarding the first question, Dennett's (2002, p. 190) bald claim that we do not know whether vision uses depictive representations is a vast overstatement. The neuroanatomy reveals that these areas play a role in the earlier phases of visual processing, and neurological data indicate that these areas play a special role in the parsing of shapes into their component parts. For example, apperceptive agnosia results when there is diffuse neural death in these areas (usually due to carbon monoxide poisoning), which in turn characteristically disrupts the ability to organize input into objects and their parts (Vecera & Gilds, 1998).

The mere fact that, as noted in previous chapters, local damage to topographic areas produces localized scotoma demonstrates that the topographic properties of these areas play a role in vision. Moreover, studies in the macaque monkey, such as those by Roelfsema and Spekreijse (2001), clearly demonstrate that the visual topography of area V1 has a functional role in visual cognition. In this and related studies, monkeys had to trace a curve mentally (out of a set of two), without moving their eyes. During a random subset of trials, the monkey made a mistake and traced the wrong curve. These mistakes could be predicted by looking at the pattern of neural activation corresponding to the topographic representation of the two curves!

Certainly, we do not yet know everything there is to know about the functioning of topographically organized visual brain areas—but this is not necessary for present purposes. The depictive properties of such areas clearly play a role in visual perception.

Regarding the second question, what about the issue of how information from successive fixations is accumulated? Ingle (2002) notes that every time the eyes move, the material in the visual buffer (i.e., the set of topographically organized areas in the occipital lobe) is written over. Therefore, the buffer itself cannot integrate information from successive fixations. This fact is, however, completely irrelevant to the issue of whether visual processing exploits depictive representations. That said, we note that every time one's eyes move, the location of the stimulus that gives rise to the perceptual image is registered in the posterior parietal lobes, which keep track of where in space one is looking when a particular part of the field is encoded. Thus, one can store successive images (presumably in the temporal lobes) plus the loci of corresponding objects (in the parietal lobes), and the two sorts of information can be integrated downstream (as we discuss in chapter 5). Moreover, the parietal representation accounts for how we see three-dimensional images when two-dimensional depictive representations are used in vision; based on a variety of cues (such as retinal disparity, texture gradients, and occlusion) processed in earlier areas, the parietal lobes index each location in space with a distance. This representation allows one later to locate and fixate again on a particular object, if needed.

## The Necessity of Interpretation

Is nothing in fact accomplished by "turning one topographically organized pattern into another"? To the contrary, there are good reasons for such mappings. For example, although the retina and area 17 have a similar spatial organization, the pattern of connections among neurons is radically different, because different information is processed in each structure. Indeed, area 17 not only has numerous short inhibitory connections, but also has neurons that compare input from two eyes (which obviously cannot be accomplished in a single retina); such information is important for segregating figure from ground and organizing perceptual units—which are not the tasks of the retina. A series of topographically mapped areas is computationally useful because their spatial organization makes explicit and accessible information needed to accomplish the initial phases of visual processing, and each phase takes advantage of transformations accomplished in earlier phases.

The fact that visual processing is not accomplished in one fell swoop was stressed by Marr (1982); vision is accomplished in a succession of small steps. We now know that visual input is processed in a set of successive brain areas, each of which abstracts specific types of information from processes that operate earlier in the sequence (e.g., Miyashita & Hayashi, 2000). The initial visual areas in cortex are topographically organized, but progressively later ones are increasingly less well organized topographically, until the latest ones in the temporal lobe no longer preserve the organization of the retina, or do so very loosely (Malach, Levy, & Hasson, 2002). Although not all visual processing relies on depictive representations, the evidence strongly indicates that depictive representations are in fact used during the early phases of visual processing.

Dennett's (2002) quip that electricity generated from hydroelectric plants is not wetter than electricity generated by other means is cute, but misses the mark. The topographical organization of early visual areas is not irrelevant in the end, after input has been fully processed. Dennett's notion is a little like saying that gasoline would only be important for automobiles if they moved ahead on a path of flaming fluids! Gasoline plays a key role in the sequence of events that allows an automobile to move, even if in the end it does not directly turn the wheels. Moreover, many features of an automobile, such as room for a gas tank and placement of an exhaust pipe so as not to allow fumes into the passenger compartment, emerge from use of an internal combustion engine. Similarly, the depictive properties of topographically organized areas play crucial roles in the early phases of visual perception, allowing the process to begin by making the spatial structure of an object explicit and accessible (with respect to processes that will operate on this representation). Furthermore, there is much evidence that the visual system is wired to allow information that is only implicit in the higher areas to be made explicit and accessible by reconstructing the shape in earlier areas. If so, then the stored representations are formed in part to have the ability to recreate such earlier, depictive representations—and this capacity is one factor that determines how the stored representations are organized.