14

Representing, perceiving, and remembering the shape of visual space

AUDE OLIVA, SOOJIN PARK, AND
TALIA KONKLE

14.1 Introduction

Our ability to recognize the current environment determines our ability to act strategically, for example when selecting a route for walking, anticipating where objects are likely to appear, and knowing what behaviors are appropriate in a particular context.

Whereas objects are typically entities that we act upon, environments are entities that we act within or navigate towards: they extend in space and encompass the observer. Because of this, we often acquire information about our surroundings by moving our head and eyes, getting at each instant a different snapshot or view of the world. Perceived snapshots are integrated with the memory of what has just been seen (Hochberg, 1986; Hollingworth and Henderson, 2004; Irwin *et al.*, 1990; Oliva *et al.*, 2004; Park and Chun, 2009), and with what has been stored over a lifetime of visual experience with the world.

In this chapter, we review studies in the behavioral, computational, and cognitive neuroscience domains that describe the role of the shape of the space in human visual perception. In other words, how do people perceive, represent, and remember the size, geometric structure, and shape features of visual scenes? One important caveat is that we typically experience space in a three-dimensional physical world, but we often study our perception of space through two-dimensional pictures. While there are likely to be important differences between the perception of space in the world and the perception of space

Vision in 3D Environments, ed. L. R. Harris and M. Jenkin. Published by Cambridge University Press.

mediated through pictures, we choose to describe in this chapter principles that are likely to apply to both media. In the following sections, we begin by describing how the properties of space can be formalized, and to what extent they influence the function and meaning of a scene. Next, we describe cases in which the perception of the geometry of space is distorted by low- and high-level influences. Then, we review studies that have examined how the memory of scenes and of position in space is transformed. Finally, we address how people get a sense of the space just beyond the view they perceive, with a review of studies on scene integration.

The visual perception of space is, first and foremost, observer-centered: the observer stands at a specific location in space, determined by latitude, longitude, and height coordinates. A *view* or *viewpoint* is a cone of visible space as seen from an observer's vantage point: a view is oriented (e.g., looking up, straight ahead, or down) and has an aperture that the dioptrics of the eyes suggest covers up to 180°. However, the apparent visual field that human observers visually experience is closer to 90°, corresponding to a hemisphere of space in front of them (Koenderink *et al.*, 2009; Pirenne, 1970). These truncated views provide the inputs provided to the brain. All ensuing spatial concepts such as scenes, places, environments, routes, and maps are constructed out of successive views of the world.

In this chapter, we introduce two levels of description of environmental spaces: a structural level and a semantic level. The terms *space*, *isovist*, and *spatial envelope* refer to the geometric context of the physical world (structural level); *scenes*, *places*, and *environments* rely on understanding the meaning of the space that the observer is looking at or embedded in (semantic level).

Whereas space is defined in physics as the opposite of mass, in our structural-level description we define a space as an entity composed of two substances: mass and holes. A space can be of any physical size in the world, e.g., 1 m^3 or 1000 m^3. The spatial arrangement of mass and holes is the most simplified version of the three-dimensional layout of the space. From a given viewpoint, the observer has access to a collection of visible surfaces between the holes. The set of surfaces visible from that location if the observer rotates through 360° is called an isovist (Benedikt, 1979). An example of an isovist is shown in Figure 14.1. A collection of all isovists visible from all possible locations in a space defines a complete isovist map of the space. One final structural-level description of the spatial layout, as seen from one viewpoint, is the spatial-envelope representation (Oliva and Torralba, 2001). Here, three-dimensional spatial layouts correspond to two-dimensional projections that can be described by a statistical representation of the image features. This statistical

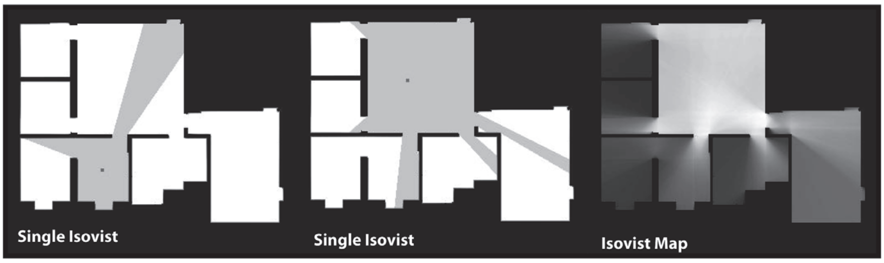


Figure 14.1 Two single isovist views are shown, with the red dot marking the location from which the isovist was generated. The rightmost image shows the isovist map, which is the collection of isovists generated from all possible locations within the space. A color version of this figure can be found on the publisher's website (www.cambridge.org/978107001756).

representation can describe coarsely the shape, size, boundary, and content of the space in view.

At the semantic level, *scene*, *place*, and *environment* are terms that refer to the meaning of the physical or pictorial world and are modulated by the knowledge of the observer. In the world, the observer is embedded in a space of a given place: a place is associated with certain actions and knowledge about a specific physical space (e.g., my kitchen or the White House) or groups of physical spaces (e.g., the category of industrial kitchens or of gymnasiums).[1] The term "scene" has two common usages in the literature, as both a particular view and an extended space. Here we define a scene as a view (or cone of visible space) with an associated semantic meaning. A scene has a "gist" (Friedman, 1979; Oliva, 2005; Potter, 1976), namely a semantic description that comes with associated knowledge (e.g., a kitchen is a place for cooking). A scene depends on one's view of a space (unlike places, which do not depend on the viewpoint of the observer). Therefore, a place can be composed of one or many scenes: by moving his or her head or moving around a city block, an observer may perceive a shop front, a parking lot, a street, and a park as different scenes. Places and scenes can be conceptualized as part of a larger topology, an environment. Environments

[1] It is important to note that the word "place" has acquired different definitions depending on the domain of study. For instance, *place* has been used interchangeably with *scene* in cognitive neuroscience (e.g., Epstein, 2008) when referring to the parahippocampal place area, or PPA. In neuroscience, the term refers to *place-cells*, which are hippocampus neurons that fire when an animal is at a particular three-dimensional location (e.g., O'Keefe, 1979). Place-cells are specified by latitude, longitude, and height coordinates, and can also be oriented, for example pointing north with a 30° downward angle.

would therefore typically refer to physical spaces encompassing one or more scenes and places, typically of a larger scale than a single place.

14.2 Representing the shape of a space

In the following sections, we describe two representations of the structure of space: the isovist representation (Benedikt, 1979) and the spatial-envelope representation (Oliva and Torralba, 2001). Both offer a formal quantitative description of how to represent a space, i.e., the volumetric structure of a scene, place, or environment. The isovist description operates over a three-dimensional model of the environment, and captures information about the distribution and arrangement of visible surfaces. The spatial-envelope description operates over projections of space onto two-dimensional views, and captures information about both the layout and the texture of surfaces.

14.2.1 Isovist representation

Figure 14.1 illustrates the isovist of a laboratory space for a given position in the center of the main room. An isovist represents the volume of space visible from a specific location, as if illuminated by a source of light at this position. As such, the isovist is observer-centered but viewpoint-independent. It represents the visible regions of space, or the shape of the place, at a given location, obtained from the observer rotating through 360°.

A concept initially introduced by Tandy (1967), the *isovist* was formalized by Benedikt (1979). Although Benedikt described an isovist as the volume visible from a given location, in a view-independent fashion, the concept can be simplified by considering a horizontal slice of the "isovist polyhedron" as illustrated by the single isovists shown in Figure 14.1. The volumetric configuration of a place requires calculating a collection of isovists at various locations: this refers to the *isovist field* or *isovist map* (Benedikt, 1979; Davis and Benedikt, 1979), shown in Figure 14.1 on the right. High luminance levels indicate areas that can be seen from most of the locations in the main central room of the laboratory and dark areas indicate regions that are hidden from most of the locations. In empty, convex rooms (such as a circular, square, or rectangular room), the isovist field is homogeneous, as every isovist from every location has the same shape and volume (or the same area if a two-dimensional floor plan is considered).

The shape of an isovist can be characterized by a set of geometrical measurements (Benedikt, 1979; Benedikt and Burnham, 1985): its *area*, corresponding to how much space can be seen from a given location; its *perimeter length*, which measures how many surfaces[2] can be seen from that location; its *variance*, which describes the degree of dispersion of the perimeter relative to the original location; and its *skewness*, which describes the asymmetry of this dispersion. All of these inform the degree to which the isovist polygon is dispersed or compact. Additional quantitative measurements of isovists have included the number of vertices (i.e., the intersections of the outlines of the isovist polygon) and the openness of the polygon. The openness of an isovist is calculated as the ratio between the length of open edges (generated by occlusions) and the length of closed edges (defined by solid visible boundaries (Psarra and Gradjewski, 2001; Wiener and Franz, 2005).

From simple geometrical measurements of isovists and isovist maps, higher-level properties of the space can be derived: its *occlusivity* (i.e., the depth to which surfaces inside the space overlap with each other (Benedikt, 1979),[3] its degree of *compactness* (a measure defined by a circle whose radius is equal to the mean radial length of the isovist, which indicates how much the isovist's shape resembles a circle), its degree of *spaciousness* (Stamps, 2005), and its degree of *convexity* (also referred to as *jaggedness*, calculated as the ratio between the squared perimeter of the isovist and its area; see Wiener and Franz (2005) and Turner *et al.* (2001)). A concave or "jagged" isovist has dents, which means that regions of the place are hidden from view. A circular, convex isovist has no hidden regions.

Our understanding of the relationship between geometrical measurements of the isovists and the perception of a scene and a place remains in its infancy. Wiener and Franz (2005) found that the degree of convexity and the openness ratio of isovists correlated with observers' judgment of the complexity of a space, which in turn, modulates navigation performance in a virtual reality environment. Simple isovist descriptors (area, occluded perimeter, variance, and skewness) predict people's impressions of the spaciousness of hotel lobbies (Benedikt and Burnham, 1985) and the degree of perceived enclosure of a room or an urban place (Stamps, 2005). Potentially, the perceptual and cognitive factors correlated with isovists and their configuration may be diagnostic of a given type of place or of the function of a space. Furthermore, behavior in a space may be predicted by these structural spatial descriptors. Along these

[2] In his 1979 paper, Benedikt defined a visible real surface as an "opaque, material, visible surface" able to scatter visible light. This disqualifies the sky, glass, mirrors, mist, and "perfectly black surfaces." Opaque boundaries are barriers that impede vision beyond them.

[3] Occlusivity measures "the length of the nonvisible radial components separating the visible space from the space one cannot see from the original location X" (Benedikt, 1979).

lines, Turner and colleagues were able to predict complex social behaviors such as way-finding and the movement of a crowd in a complex environment (Turner *et al.*, 2001).

An analysis of the kinds of space that a human being encounters, and of the geometrical properties that distinguish different kinds of spaces from each other remains to be done. Furthermore, in its original form, the isovist theory does not account for the types of textures, materials, or colors attached to the surfaces, and this information will likely be important for relating structural descriptions of spaces to human perceptions of spaces or actions within spaces. However, the isovist description does provide a global geometrical analysis of the spatial environment and gives mathematical descriptions to spatial terms such as "vista", "panorama," and "enclosure", which in turn allows us to formalize and predict spatial behaviors of human, animal, and artificial systems. In the next section, we describe another formal approach for describing the shape of a space, the spatial-envelope representation.

14.2.2 *Spatial-envelope representation*

Given that we experience a three-dimensional world, it makes sense that we have learned to associate the meaning of a scene with properties that are diagnostic of the spatial layout and geometry, as well as with the objects in view (e.g., while closets typically contain clothes, and gyms typically contain exercise equipment, it is also the case that closets are typically very small places and gyms are large places).

In architecture, the term "spatial envelope" refers to a description of a whole space that provides an "instant impression of the volume of a room or an urban site" (Michel, 1996). The concept has been used to describe qualitatively the character and mood of a physical or pictorial space, represented by its boundaries (e.g., walls, floor, ceiling, and lighting) stripped of movable elements (e.g., objects and furnishing).

In 2001, Oliva and Torralba extended this concept and proposed a formal, computational approach to the capture of the shape of space as it would be perceived from an observer's vantage point (Oliva and Torralba, 2001, 2002, 2006, 2007; Torralba and Oliva, 2002, 2003). The collection of properties describing a space in view is referred to as the spatial-envelope representation. For instance, just as a face can be described by attributes such as its size, gender, age, symmetry, emotion, attractiveness, skin type, or configuration of facial features, a space can be described by a collection of properties such as perspective, size, dominant depth, openness, and naturalness of content.

To give an example of these scene properties, a space can be represented by two independent descriptors, one representing the boundaries or external

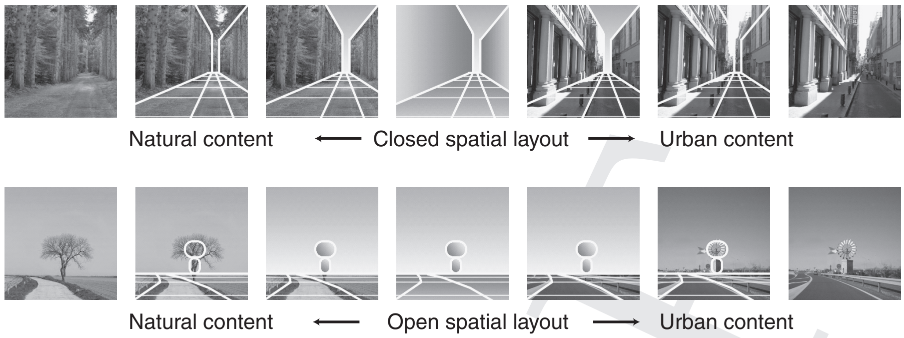


Figure 14.2 A schematic illustration of how pictures of real-world scenes can be defined uniquely by their spatial layout and content. Note that the configuration, size, and locations of components can be in correspondence between natural and manufactured environments. If we strip off the natural content of a forest, keeping the enclosed spatial layout, and fill the space with urban content, then the scene becomes an urban street scene. If we strip off the natural content of a field, keeping the open spatial layout, and fill the space with urban content, then the scene becomes an urban highway. A color version of this figure can be found on the publisher's website (www.cambridge.org/978107001756).

features, and one representing the content or internal features (Oliva and Torralba, 2001, 2002, 2006; Park *et al.*, 2011). Boundaries and content descriptors are orthogonal properties: a space can be of various sizes and shapes, and it can have any content in terms of parts, textures, colors, and materials. Figure 14.2 illustrates this point: a space can have either a closed or an open layout of a particular shape (the enclosed layout here is in perspective, and the open layout has a central figure), with its surface boundaries "painted" with either natural or manufactured content.

Oliva and Torralba (2001) discovered that some of the key properties of the spatial envelope (e.g., mean depth, openness, perspective, naturalness, and roughness) have a direct transposition into visual features of two-dimensional surfaces. This allows the calculation of the degree of openness, perspective, mean depth, or naturalness of a scene by capturing the distribution of local image features and determining the visual elements (oriented contours, spatial frequencies, and spatial resolution) diagnostic of a particular spatial layout (Oliva and Torralba 2001; Torralba and Oliva, 2002; Ross and Oliva, 2010). This statistical representation of the spatial distribution of local image features is compressed relative to the original image. To visualize what information is contained in this spatial-envelope representation, sketch images are shown in Figure 14.3 below the original image, where random noise was coerced to have

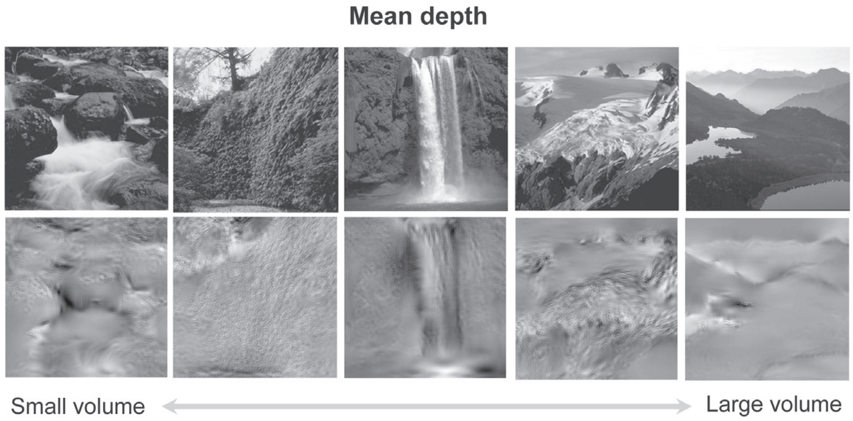


Figure 14.3 Top: examples of natural-scene images with different degrees of mean depth (from small to large volume). Bottom: a sketch representation of the visual features captured with the spatial-envelope representation (see Oliva and Torralba (2001) and (2006) for details). Note that this representation of a natural scene has no explicit coding of objects or segmented regions. A color version of this figure can be found on the publisher's website (www.cambridge.org/978107001756).

the same statistical representation as the original image (see Oliva and Torralba (2006) for details).

A summary of the framework of the spatial-envelope model is shown in Figure 14.4. For simplicity, the model is presented here as a combination of four global scene properties (Figure 14.4a). The implementation of the model takes the form of high-level image filters originating from the outputs of local oriented filters, as in the early visual areas of the brain (Figure 14.4b). Within this framework, the structure of a scene is characterized by the properties of the boundaries of the space (e.g., the size of the space, its degree of openness, and the perspective) and the properties of its content (e.g., the style of the surfaces, whether the scene is natural or artificial, the roughness of these surfaces, the level of clutter, and the type of materials). Any scene image can be described by the values it takes for each spatial-envelope property. These values can then be represented by terms that describe, for instance, the degree of openness of a given scene ("very open/panoramic", "open", "closed," or "very closed/enclosed"); (Oliva and Torralba, 2002). In this framework, instead of a forest being described as an environment with trees, bushes, and leaves, it would be described at an intermediate level as "a natural enclosed environment with a dense, isotropic texture." Similarly, a specific image of a street scene could be described as an "artificial outdoor place with perspective and a medium level of

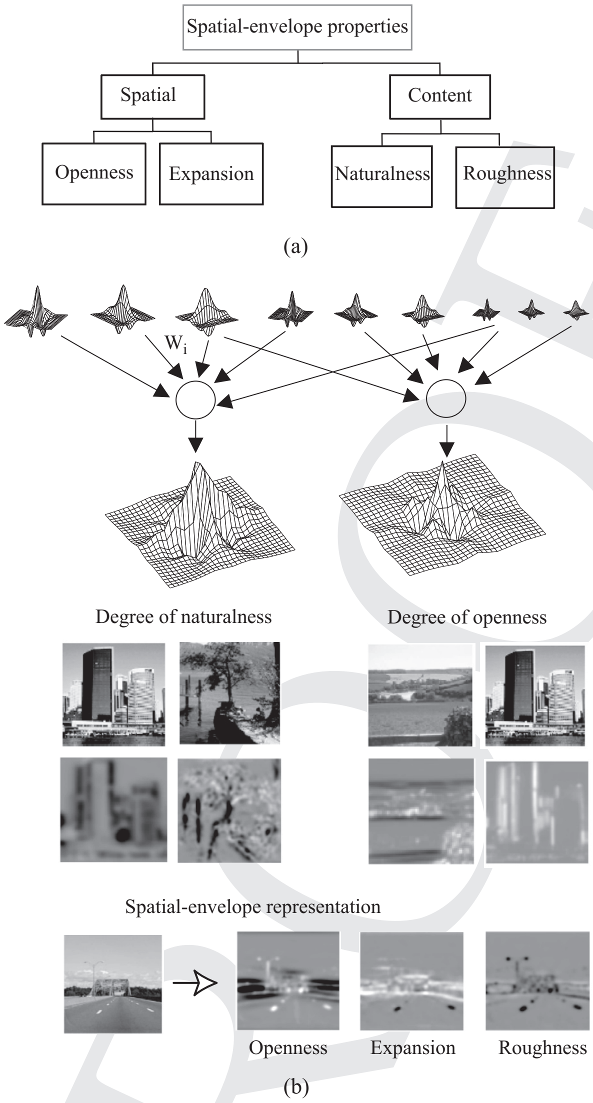
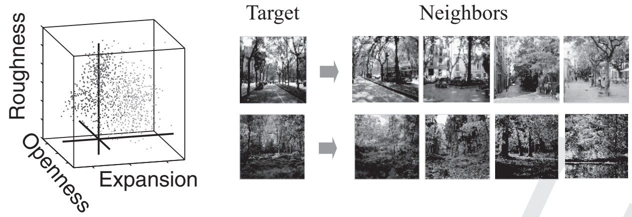


Figure 14.4 Schematic spatial-envelope model. (a) Spatial-envelope properties can be classified into spatial and content properties. (b) Illustration of a computational-neuroscience implementation of the spatial-envelope model. The features of naturalness and openness are illustrated here. (c) Projection pictures of artificial environments onto three spatial-envelope dimensions, creating a scene space (based on global properties only, no representation of objects here). Semantic categories (different colors) emerge, showing that the spatial-envelope representation carries information about the semantic class of a scene. Two target images, together with their nearest neighbors in the spatial-envelope space, are shown here (from a dense database of images). A color version of this figure can be found on the publisher's website (www.cambridge.org/978107001756).



(c)

Figure 14.4 Continued

clutter." This level of description is meaningful to observers who can infer the probable semantic category of the scene. Indeed, Oliva and Torralba observed that scene images judged by people to have the same categorical membership (street, highway, forest, coastline, etc.) were projected close together in a multidimensional space whose axes correspond to the spatial-envelope dimensions (Figure 14.4c). Neighboring images in the spatial-envelope space correspond to images with a similar spatial layout and a similar semantic description (more so when the space is filled densely, i.e., either with a lot of varied exemplars or with typical exemplars of categories).

As shown in these sections, both the isovist and the spatial-envelope representations provide many interesting and complementary descriptors of the shape of the space that are quantitatively defined. The isovist describes the visible volumes of a three-dimensional space, while the spatial envelope captures layout and content features from a two-dimensional projected view. In these theories, space is a material entity as important as any other surface, such as wood, glass, or rock. Space has a shape with external and internal parts that can be represented by algorithms and quantitative measurements, some of which are very similar to operations likely to be implemented in the brain. These approaches constitute different instances of a space-centered understanding of the world, as opposed to an object-centered approach (Barnard and Forsyth, 2001; Carson *et al.*, 2002; Marr, 1982).

14.3 Perceiving the shape of a space

Numerous studies have shown that our perception of space is not veridical: it can be distorted by a number of factors. Some factors are basic constraints arising from visual-field resolution and the challenge of recovering the three-dimensional structure from a two-dimensional projection on the retinas.

Other factors go beyond simple optics and include top-down effects of knowledge, as well as markers reflecting our physiological state. Finally, systematic distortions can arise as a consequence of perceptual dynamics as we adapt to the volumetric properties of the space around us. In this section, we will focus our review on distortions that change our global perception of the overall shape, volume, distances or slants of a space, rather than our local perception centered on objects or parts.

14.3.1 Distortion of the geometry of a space

There are many ways in which our perception of space is not veridical: for example, distances in the frontal plane (i.e., traversing from left to right) appear much larger than distances in the sagittal plane (i.e., receding in depth from the observer) (Wagner, 1985; Loomis, *et al.*, 1992), while distances in the frontal plane appear much smaller than vertical distances (e.g., Higashiyama, 1996; Yang *et al.*, 1999). Surface angles are often underestimated and slants of hills are often overestimated (Proffitt *et al.*, 1995; Creem-Regehr *et al.*, 2004). Distances to objects can be misperceived when a relatively wide expanse of the ground surface is not visible (Wu *et al.*, 2004), or when the field of view is too narrow (Fortenbaugh *et al.*, 2007). Such biases are also highly dependent on the structure of the scene: distance judgments are most difficult and inaccurate in a corridor; they are easy, accurate, and reliable out in an open field; and they are easy and reliable but inaccurate in a lobby (Lappin *et al.*, 2006). Many visual illusions, for example the Ames room, take advantage of different depth cues to change the perception of the size of objects and the size of a space.

The rules for the distortion of the perception of physical space have been well documented (for a review, see Cutting (2003)): as physical distance increases, perceived distances are foreshortened as compared with physical space (Loomis and Philbeck, 1999). This means that observers do not accurately evaluate distances between objects at far distances, being only able to judge ordinal relations (which surface is in front of another, but not by how much). The compression of planes in a space with distance of viewing is likely to be due to the decrease in the available information and depth cues (Indow, 1991, see Cutting, 2003). In his 2003 review, Cutting reports three classes of ecological space perception ranges. First, perception in the *personal space* (up to about 2–3 m), is metric: indeed, veridical spatial computation is necessary for accurate hand reaching and grasping. In close-up space, distance to objects and surfaces are provided by many sources of information and depth cues, including accommodation and convergence, that cease to be effective beyond a few meters (Loomis *et al.*, 1996). Second, the *action space* is defined in practice by the distance to which one can throw an object accurately (up to about 30 m away). Whereas depth perception in the action space suffers some compression, studies found it to be close to physical space. Beyond a few tens of meters is *vista space*, where an observer's perception of distances to and between surfaces can become greatly inaccurate, with a dramatically accelerating foreshortening of space perception for distances over 100 m. At that range, traditional pictorial cues of information are in effect (e.g., occlusion, relative size, aerial perspective, height in the visual field, and relative density; see Cutting, 2003). Observers rely on thier knowledge of the relative sizes of objects, and ordinal cues such as layout arrangement and occlusion to infer the shape of the three-dimensional space. Furthermore, these drastic spatial compressions of vista space are not noticed by individuals (Cutting and Vishton, 1995; Cutting, 2003).

This gradient of the perceived compression of space suggests the need for a perceptual isovist, where the characteristics of the shape of visible surfaces are measured not from actual distances in a volume but from perceived distances (see Section 14.2.1). For example, we would expect a more deformed isovist for background than for foreground planes, since perception is based on ordinal estimations for the background planes. When only ordinal depth information about planes is available, some illusions of scene volume and misinterpretation of surfaces may occur. Figure 14.5 illustrates these illusions using photographs of natural scenes: the "mountain cliff" and the picture of a "river receding into the distance" (Figures 14.5a,c) are perceived as "the base of a mountain" and "a view looking up at the sky" (Figures 14.5b,d), respectively, when the images are inverted. Here, the image inversion has two main effects: it reverses lighting effects, which may change the surfaces' affiliation as "object" and "ground," and in some cases it produces large changes in the perceived scale of the space. The spatial-envelope approach (Oliva and Torralba, 2001; Torralba and Oliva,

Figure 14.5 Examples of natural images in which inverting the images creates a plausible scene view with dramatic changes in the interpretation of surfaces and volume between upright and upside down (adapted from Torralba and Oliva, 2003). A color version of this figure can be found on the publisher's website (www.cambridge.org/978107001756).

2002) captures the low-level and texture statistics which are correlated with the change of perceived scale and semantics.

14.3.2 *Changing the volume of a space*

The tilt–shift illusion is another scene depth illusion where a small change in the levels of blur across an image can make an expansive scene look miniature (Figure 14.6). The degree of focus across a scene is a simple low-level depth cue: for example, as you fixate out at more distant points in space, the angle between your eyes narrows (accommodation), which influences the retinal blur gradient (Watt *et al.*, 2005; Held and Banks, 2008). For example, focusing on an object very close in front of you will lead to a situation where only a small portion of the image can be in focus, with the upper and lower parts of the scene blurred. Thus expansive scenes can be made to look small by adding blur. This effect works best for scenes which are taken from high above, mimicking the angle of view one would have if looking at a toy model. In other words, the tilt–shift effect works by changing the low-level statistics (the blur and angle of surfaces caused by an elevated head angle) to influence the perceived volume of a space.

While the tilt–shift makes a large scene look small, the converse is also possible. Making small scenes look large is a trick that has been honed to an art by Hollywood special effects artists. The original special effects took advantage of two-dimensional projection rules, in a technique called "forced perspective." For example, suppose cars are traveling across a bridge, with a camera filming the scene from afar. By putting a model version of the bridge much closer to the

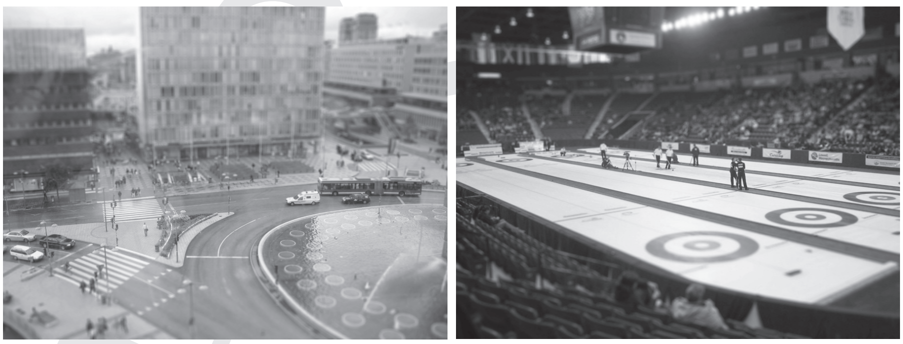


Figure 14.6 Two examples of the tilt–shift illusion, where adding a blur gradient to the upper and lower portions of an image makes the scene appear miniature. A color version of this figure can be found on the publisher's website (www.cambridge.org/978107001756).

camera, the real bridge and the model bridge can be set to project to the same two-dimensional image, allowing a dramatic explosion of the model bridge to look real.

These are examples of depth illusions, where the volume of the space changes based on the cues in the environment and on our expectations about the structure and statistics of the natural world. Indeed, neuroimaging work has shown that the size that you think something is in the world matters beyond just the visual angle at which it projects onto the retina. Murray *et al.* (2006), presented observers with two disks with matched visual angles on the screen, but with contextual information that made one disk look much larger (and farther away) than the other. The bigger disk activated a greater extent of the primary visual cortex than the smaller disk did, despite their equivalent visual size. These results suggest that the perceived physical size of an object for space has consequences on very early stages of visual processing.

14.3.3 *Changing the percept of a space: top-down influences*

Distance estimation, like time, is modulated by individuals' subjective perception: buying two gallons of milk instead of a carton of cream to carry back from the grocery store can make you feel that you are further from home. Interestingly, work by Proffitt and collaborators (Proffitt *et al.*, 2003) shows that nonoptical cognitive variables may influence the perception of space cues.

Along with task constraints, the physical resources and capabilities available to an agent change the perception of space (e.g., distance and slant angle of ground surfaces). For example, participants wearing heavy backpacks thought that a target object on the ground was located further away from the starting point than did individuals who were not wearing backpacks (Proffitt *et al.*, 2003). Importantly, such modulation of distance estimation occurred only when the participants intended to walk the distance (Witt *et al.*, 2004). Similarly, the manipulation with a backpack load had no effect on people's distance estimation when they were asked to throw a ball in the direction of the object. However, the weight of the ball changed the estimation of distance for participants who intended to throw the ball. In other words, only when the increased physical effort was directly related to the intended action did the estimation of the distance change (Witt *et al.*, 2004; but see also Woods *et al.*, 2009).

Other studies have shown that the inherent characteristics or physical capabilities of an individual can also influence how they perceive space. For example, compared with younger people, older people with low physical capabilities tend to estimate a distance as longer or the same hill as steeper (Bhalla and Proffitt, 1999). When younger participants are primed with an elderly stereotype, they also have a tendency to overestimate distances (Twedt *et al.*, 2009). The psychosocial state of an individual might also influence the perception of the space. Proffitt and colleagues found that participants who imagined positive social contact estimated the slant of a hill to be less than did participants who imagined neutral or negative social contact (Schnall *et al.*, 2008). Although it is hard to conclude from these studies whether these nonoptical factors fundamentally changed the observers' perceptions or whether they modulated responses without changing perception, they provide evidence that the experience of the geographical properties of a space can be influenced by changes in the psychological load of an observer, beyond the attributes of the physical world.

14.3.4 Adaptation to spatial layout

The previous sections have presented examples of how spatial low-level cues and preexisting top-down knowledge can influence our perception of the space that we are looking at, even if our physical view of the world stays the same. Similarly, temporal history can also influence the perception of a space: the experience that you had with particular visual scenes just moments ago can change your perception of the structure and depth of a scene that you are currently viewing. This is the phenomenon of adaptation: if observers are overexposed to certain visual features, adaptation to those features affects the conscious perception of a subsequently presented stimulus (for example, this is classically demonstrated by adapting to a grating moving in one direction, where, afterwards, a static grating will appear to move in the opposite direction).

Using an adaptation paradigm, Greene and Oliva (2010) tested whether observers adapt to global spatial-envelope properties (described in Section 14.2.2), such as mean depth and openness. In one study, observers were presented with a stream of natural scenes which were largely different (in terms of categories, colors, layout, etc.), but which were all exemplars of very open scenes, representing vista space (panorama views of fields, coastlines, deserts, beaches, mountains, etc.). Following this adaptation phase, a scene picture with a medium level of openness (e.g., a landscape with a background element) was presented for a short duration, and observers had to quickly decide whether this scene was very open or very closed. When observers were adapted to a stream of open scenes, ambiguous test images were more likely to be judged as closed. In contrast, the same ambiguous test images were judged to be open following adaptation to a stream of closed scenes (e.g., caves, forests, or canyons). Similar aftereffects occurred after observers had adapted to other extremes of spatial-envelope properties, such as small versus large depth and natural versus urban spaces, and even to higher-level properties of the scene, such as when the view depicted an environment with a hot versus a cold climate.

Figure 14.7 Continuum between forests and fields. Images in the middle of the continuum have an ambiguous category and can be perceived as both a field and as a forest. A color version of this figure can be found on the publisher's website (www.cambridge.org/978107001756).

Importantly, Greene and Oliva showed further that adaptation to different scene envelope properties not only influenced judgments of the corresponding scene properties for a new image but also influenced categorical judgments about a new image. This experiment took advantage of the fact that fields are usually open scenes, while forests are typically closed scenes. Importantly, there is a continuum between field and forest scenes, with some scenes existing ambiguously between the two categories that can be perceived either as a field or as a forest (see Figure 14.7).

During the adaptation phase, observers were presented with a stream of images which again varied in their basic level category and their surface features, but all depicted open or closed views. No forests or fields were presented in this stream of images. After the observers had adapted to open scenes, an ambiguously open or closed image would be expected appear to be more closed. The critical question was whether an ambiguous field/forest image would also be more likely to be judged as a forest than as a field, which has a more enclosed property. Similarly, adapting to a stream of closed natural images should cause the same ambiguous field/forest to be more likely to be judged as a field. Indeed, this is exactly what Greene and Oliva observed.

These results demonstrate that exposure to a variety of scenes with a shared spatial property can influence the observers' judgments of that spatial property later, and can even influence the semantic categorization of a scene. Such adaptation aftereffects have been shown for low-level features such as orientation and motion (Wade and Verstraten, 2005), and even high-level features such as shape, face identity, and gender (Leopold *et al.*, 2001; Webster *et al.*, 2004). The adaptation mechanisms suggest that the neural system is tracking the statistics of the visual input and tuning its response properties to match. Thus, the aftereffects for global scene properties broadly imply that as observers process natural scenes, one of the extracted statistics to which the system is tuned reflects the layout and perceived volume of the scene.

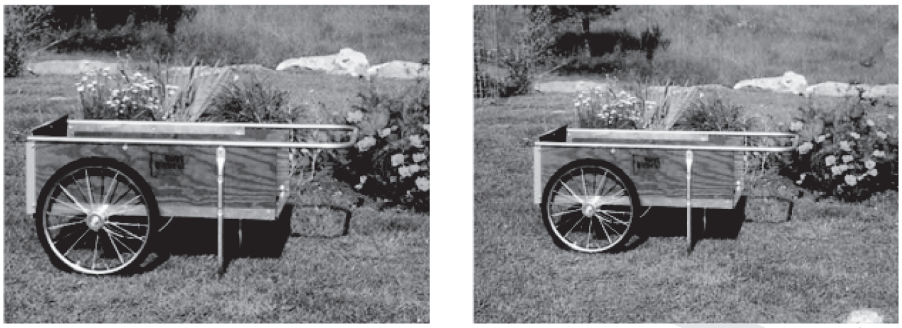
14.4 Remembering the shape of a space

The previous section has reviewed evidence that the perception of space can be manipulated by low-level image cues, top-down influences, and the temporal history of scenes. These perceptual illusions occur online while the relevant sensory information is present in the world, but similar systematic distortions of space occur when we represent scene information that is no longer in view but is instead held in memory. In the following sections, we discuss how single views are remembered and how this effect might be understood in the framework of navigation through a space.

14.4.1 Behavioral and neural aspects of boundary extension

When presented with a scene view, what do observers remember about the space depicted? Intraub and Richardson (1989) presented observers with pictures of scenes, and found that when the observers drew the scenes from memory, they systematically drew more of the space than was actually shown: this is the phenomenon of boundary extension. Since this initial demonstration, much research has been done showing the generality of this effect. For example, boundary extension is robust to various tasks beyond drawing, such as rating and border adjustment (e.g., Intraub *et al.*, 1992, 2006), and to different image sets (Candel *et al.*, 2003; Daniels and Intraub, 2006); it operates over a range of timescales from minutes to hours (Intraub and Dickinson, 2008); and it is found in both young children and older adults (Candel *et al.*, 2004; Seamon *et al.*, 2002). Interestingly, boundary extension occurs even when observers are blindfolded – they explore space with their hands – suggesting an important link between the representations of space across sensory modalities (Intraub, 2004). Figure 14.8 shows an example of boundary extension. Observers presented with the scene in Figure 14.8a will remember the scene as having more information around the edges, as depicted in Figure 14.8b.

In a functional neuroimaging study, Park *et al.* (2007) examined whether scene-selective neural regions showed evidence of representing more space than the original scene view. Critically, they used a neural adaptation paradigm (also called repetition attenuation); (Grill-Spector *et al.*, 2005) to determine what scene information was being represented. In an adaptation paradigm, when a stimulus is repeated, the amount of neural activity is reduced when processed for the second time compared with when it was processed as a novel stimulus. This logic suggests that a second presentation of the stimulus matches what



(a) (b)

Figure 14.8 Example of boundary extension. After viewing a close-up view of a scene (a), observers tend to report an extended representation (b). A color version of this figure can be found on the publisher's website (www.cambridge.org/ 978107001756).

was previously presented, thereby facilitating visual processing and reducing neural activity. Park and colleagues used the phenomenon of neural adaptation to examine whether the brain's sensitivity to scene views was consistent with predictions derived from the phenomenon of boundary extension. When an observer is presented with a close scene view, the existence of boundary extension predicts that this scene view might be represented at a wider angle than that at which it was originally presented. Thus, if the second stimulus is presented slightly wider than the original, this should match the representation in scene-selective areas and show a large degree of attenuation. Conversely, if the order of these stimuli is reversed, the representation of the wide-angle view will be very different from that of a subsequently presented close view, and thus no neural attention is expected. This is precisely the pattern of results that Park *et al.* (2007) observed in the parahippocampal place area, as shown in Figure 14.9.

14.4.2 Navigating to remembered scene views

While boundary extension can be interpreted as an extrapolation of information in the periphery of a scene (requiring no movement of the observer), this effect can also be examined within a three-dimensional environment. Here we explore the notion of a prototypical view and examine whether the memory of a view from a specific location might be influenced by a prototypical view.

In general, a view of a scene arises from an observer's location in a three-dimensional space. As the observer walks through an environment, the view

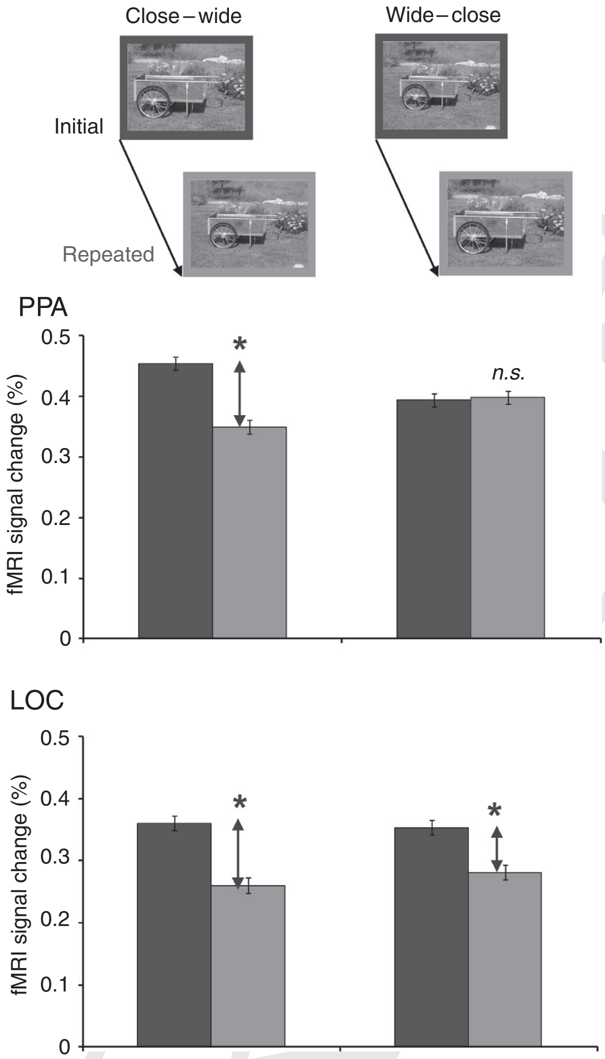


Figure 14.9 Examples of close–wide and wide–close conditions are presented in the top row. The peaks of the hemodynamic responses for close–wide and wide–close conditions are shown for the PPA and LOC. An interaction between the activations for the close–wide and wide–close conditions representing boundary extension asymmetry, was observed in the PPA but not in the LOC. The error bars indicate the standard error (± s.e.m.). Figure adapted from Park *et al.* (2007). A color version of this figure can be found on the publisher's website (www.cambridge.org/ 978107001756).

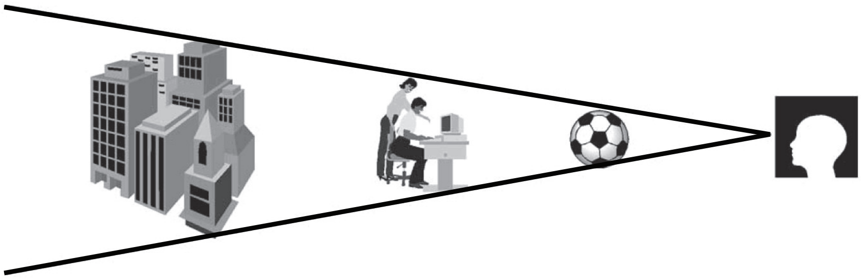


Figure 14.10 The semantic meaning of a scene changes as the depth of the scene increases. From a large distance, an observer may view buildings, which as the observer approaches change to rooms or to singleton objects on a surface. Figure adapted from Torralba and Oliva (2002).

gives rise to a scene gist (e.g., a forest) that changes slowly as the observer walks forward (e.g., a house view, followed by a view of a foyer, then a corridor, and then a bedroom). In other words, different views may take on new semantic interpretations at different spatial scales (Oliva and Torralba, 2001; Torralba and Oliva, 2002) (Figure 14.10). However, there also are many views with the same scene gist (e.g., a bedroom), which remain consistent whether the observer walks a few steps backwards or a few steps forward. Given these different views of a scene, is there a prototypical location within a volume that gives rise to a consistently preferred scene view?

Konkle and Oliva (2007a) examined this question by placing observers at either the front or the back of a "virtual room" and had them maneuver forward or backward through the space until they had the best view. We did not define what the "best view" was for observers, but provided people with instructions reminiscent of the story of Goldilocks and the three bears: "this very close view is too close, and this very far view is too far, so somewhere in between is a view that is just right." In all our rooms, the three-dimensional space was constructed so that all locations and views had the same semantic gist of the scene. Two places are shown in Figure 14.11, which shows the closest possible view (left), the farthest possible view (right), and the preferred view across all observers (middle).

Despite the subjectivity of the task, the observers were relatively consistent in their preferred views, and most used a consistent navigation strategy in which they moved all the way to the back of the scene, for example "I zoomed out to see what type of space it was," and then walked forward "until I felt comfortable "/" until it looked right." A few observers commented that to get the best view they

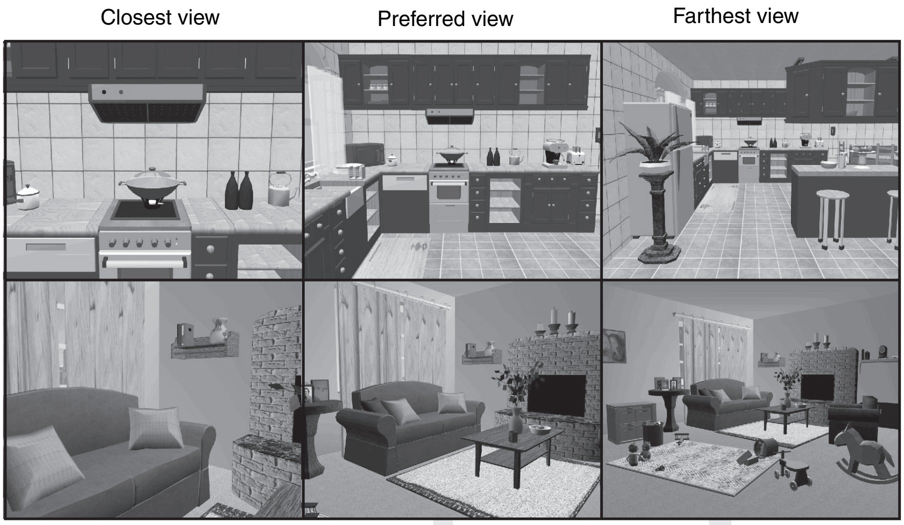


Figure 14.11 Two example spaces, a kitchen (top) and a living room (bottom). The closest and farthest views of the scenes are shown (left and right, respectively), as well as the preferred scene view across observers (middle). A color version of this figure can be found on the publisher's website (www.cambridge.org/978107001756).

wanted to step either left or right, which was not allowed in the experimental design. The data suggest that given a scene, some views are indeed better than others, and observers have a sense of how to walk to get the best view. In geometric terms, this notion that there is a prototypical view implies that there is a particular preferred viewing location in 3D space.

Konkle and Oliva (2007a) next tested memory for scene views along the walking path (from the entrance view to the close-up view). Observers studied particular scene views for each of the rooms, where some views were close up and others were wide angle, defined relative to the preferred view. To test memory for these scene views, the observers were placed in the room at either the back or the front of the space and had to maneuver through the space to match where they stood during the study phase (for a similar method, see Konkle and Oliva, 2007b).

The results showed that for the close-up views, observers tended to navigate to a position farther back in the scene, showing boundary extension. For far views, the opposite pattern was observed, where people tended to navigate to a closer location than the view studied (Figure 14.12). Thus, in this experimental task with these scenes, we observed boundary extension for close views and boundary restriction for far views. Importantly, the memory errors were not

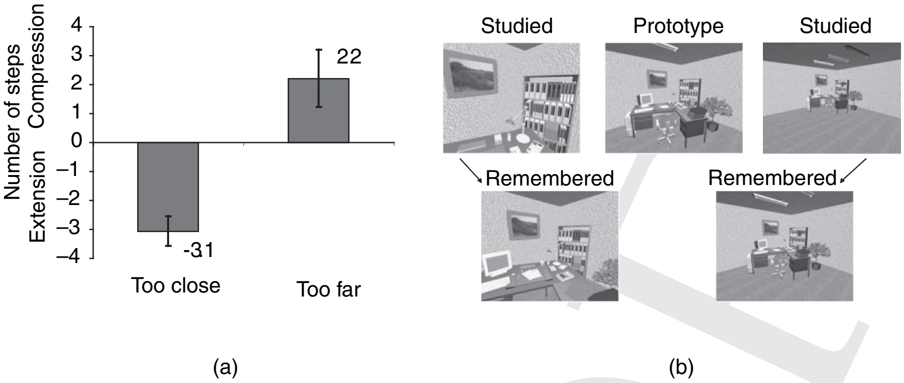


Figure 14.12 (a) Memory errors for scenes presented too close or too far, measured by number of steps. The error bars reflect ±1 s.e.m. (b) Example scene used in the memory experiment. Observers were presented with either a too-close or a too-far view. The remembered scene views were systematically biased towards the prototypical view. A color version of this figure can be found on the publisher's website (www.cambridge.org/978107001756).

driven by a few large errors (e.g., as if observers sometimes selected a very far scene rather than a very close one), but instead reflect small shifts of one or two virtual steps. While boundary restriction is not often observed, one possible explanation for why we observed both boundary extension and boundary restriction is that the close and far views used here cover a large range of space (the "action space", see Section 14.4.2).

These systematic biases in memory can be explained by the notion that memory for a scene is reconstructive (e.g., Bartlett, 1932). According to this idea, when an observer has to navigate to match a scene view in memory, if they have any uncertainty about the location, then they will not guess randomly from among the options, but instead will choose a view that is closer to the prototypical view. This way, the memory for a particular scene representation can take advantage of the regularities observed in other scenes of that semantic category and spatial layout to support a more robust memory trace. While this strategy will lead to small systematic memory errors towards the prototypical view, it is actually an optimal strategy to improve memory accuracy overall (Huttenlocher *et al.*, 2000; Hemmer and Steyvers, 2009).

Currently, there is still much to understand about what aspects of natural visual scenes determine the magnitude and direction of memory errors. For example, some work suggests that boundary extension errors can depend on the identity and size of the central object (e.g., Bertamini *et al.*, 2005) and on the complexity of the background scene (e.g., Gallagher *et al.*, 2005). The relative contributions of structural features of the space and semantic features of the scene in these effects are unknown, and we cannot yet take an arbitrary scene and understand what spatial distortions will be present in memory. We believe that such predictions will become possible as we gain a richer and more quantitative vocabulary that characterizes the many possible spatial relationships between an observer and the elements in front of him or her, as well as the spatial structure of the three-dimensional space.

Finally, in all of these experiments, observers have to remember a scene view that is presented at a visual angle subtending 5° to 20°. However, in natural viewing conditions, observers navigate through the environment with a full-field view. While previous studies have demonstrated that the shape of the aperture (rectilinear, oval, or irregular) does not affect the magnitude of boundary extension (Daniels and Intraub, 2006), one important question is whether these memory biases necessarily depend on a restricted view of a scene relative to our whole visual field. To test this, we had observers complete the same task with a full-field display (Figure 14.13) T. Konkle, M. Boucart, and A. Oliva, unpublished data). We found that the observers showed similar memory errors, with boundary extension in memory for close-up views, boundary compression in memory for far away views, and no systematic bias for prototypical scene views.

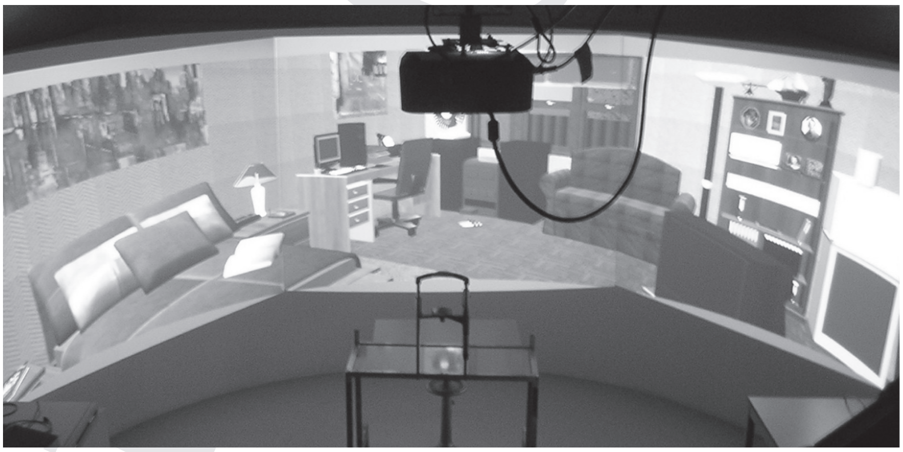


Figure 14.13 Scene view presented in a full-field display. Observers were seated at the table, with their head position fixed by the chin rest. A color version of this figure can be found on the publisher's website (www.cambridge.org/978107001756).

The data from this panoramic study demonstrate that these scene memory mechanisms discovered using pictorial scenes presented on a monitor operate even on scenes presented to the full visual field. Overall, these data support the notion that prototypical views may serve as an anchor for the memory of a specific view, and that scene-processing mechanisms may serve not only to help construct a continuous world, but also to support optimal views for perception and memory of a three-dimensional space.

14.5 From views to volume: integrating space

People experience space in a variety of ways, sometimes viewing the scene through an aperture but sometimes becoming immersed in an environment that extends beyond what can be perceived in a single view. Numerous studies have shown that the brain makes predictions about what may exist in the world beyond the aperture-like visual input by using visual associations or context (Bar, 2004; Chun, 2000; Palmer, 1975, among others), by combining the current scene with recent experience in perceptual or short-term memory (Irwin *et al.*, 1990; Lyle and Johnson, 2006; Miller and Gazzaniga, 1998; Oliva *et al.*, 2004), and by extrapolating scene boundaries (Intraub and Richardson, 1989; Hochberg, 1986) (see Section 14.4). These predictions and extrapolations help build a coherent percept of the world (Hochberg, 1978, 1986; Kanizsa and Gerbino, 1982).

Park and Chun (2009) recently tested whether the brain holds an explicit neural representation of a place beyond the scene in view. In an fMRI scanner, participants were presented with three consecutive, overlapping views from a single panoramic scene (Figure 14.14), so that the observers perceived a natural scan of the environment, as if moving their head from left to right. The researchers investigated whether brain regions known to respond preferentially to pictures of natural scenes and spaces also show sensitivity to views that are integrated into a coherent panorama.

Park and Chun (2009) found that the parahippocampal place area (PPA), an area known to represent scenes and spatial-layout properties (Epstein and Kanwisher, 1998; Park *et al.*, 2011), has a view-specific representation (see also Epstein *et al.*, 2003): the PPA treated each view of the panoramic scene as a different "scene." In contrast, the retrosplenial cortex (RSC), an area implicated in navigation and route learning in humans and rodents (Burgess *et al.*, 2001; Aguirre and D'Esposito, 1999) (see also Vann *et al.*, (2009) for a review) exhibited view-invariant representation: the RSC treated all three different views as a single continuous place, as expressed by neural attenuation from view 1

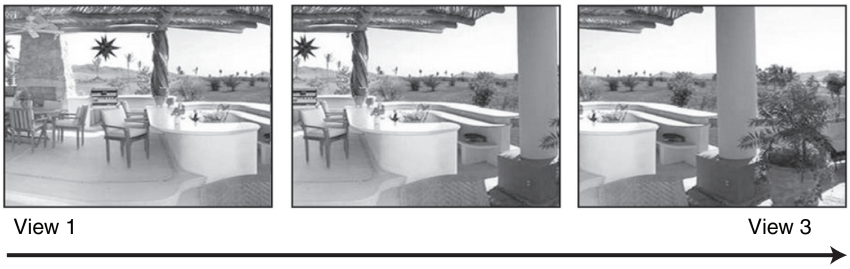


Figure 14.14 The first, and second, third views from a single panoramic scene. Views 1, 2 and 3 were sequentially presented one at a time at fixation. The first and the third view overlapped in 33% of their physical details. A color version of this figure can be found on the publisher's website (www.cambridge.org/978107001756).

to view 3. Additional experiments suggested that the RSC showed such neural attenuation only when the views were displayed in close spatiotemporal continuity. When the same trials were presented with a longer lag or with intervening items between views, the RSC no longer showed neural attenuation and responded highly to each view as if it was a novel scene. In summary, the PPA and RSC appear to complement each other by representing both view-specific and view-invariant information from scenes in a place.

While Park and Chun (2009) tested the extrapolation of views at a local level (e.g., by scanning the world through simulated head and eye movements while the viewer's location was constant), Epstein *et al.*, (2007) tested the neural basis of the extrapolation of views to a larger volume, beyond the viewer's current location. In their study, Epstein *et al.* presented participants from the University of Pennsylvania community with views of familiar places around the campus or views from a different, unfamiliar campus. The participants' tasks were to judge the location of each view (e.g., whether on the west or east of 36th Street) or its orientation (e.g., whether it was facing to the west or east of the campus). Whereas the PPA responded equally to all conditions, the RSC activation was strongest for location judgments. This task required information about the viewer's current location, as well as the location of the current scene within the larger environment. The RSC activation was second highest when viewers were making orientation judgments, which required information about the viewer's location and head direction, but not the location of the current scene relative to the environment. The RSC responded less highly in the familiar condition and the least in the unfamiliar condition. These graded modulations of RSC activity suggest that this region is strongly involved in the retrieval of long-term spatial knowledge, such as the location of a viewer within a scene, and the location of a scene within a bigger environment. The involvement of the RSC in the retrieval of long-term memory is consistent with patient and neuroimaging studies that have shown the involvement of the RSC in episodic and autobiographic memory retrieval (Burgess *et al.*, 2001; Maguire, 2001; Byrne *et al.*, 2007).

In a related vein, several, spatial navigation studies suggest that people can use geometric environmental cues such as landmarks (Burgess, 2006; McNamara *et al.*, 2003) or alignments with respect to walls to recognize a novel view of the same place as fast as a learned view, suggesting that people represent places or environments beyond the visual input. Interestingly, the modern world is full of spatial leaps and categorical continuity ruptures between scenes that violate the expectations we have about the geometrical relationship between places within a given environment. For instance, subways act like "wormholes" (Rothman and Warren, 2006; Schnapp and Warren, 2007; Ericson and Warren, 2009), distorting the perception of the spatiotemporal relationships between the locations of places in a geometrical map. Warren and colleagues tested how people behave in such "rips" and "folds" using a maze in a virtual reality world. When participants were asked to walk between two objects at different locations in a maze, they naturally took advantage of wormhole shortcuts and avoided going around a longer path. The observers did not notice that the wormholes violated the Euclidean structure of the geometrical map of the maze. These results demonstrate that the spatial knowledge about a broad environment does not exist as a complete integrated cognitive map per se, but instead exists as a combination of local neighborhood directions and distances embedded in a weak topological structure of the world.

Altogether, human spatial perception is not restricted to the current view of an aperture, but expands to the broader environmental space by representing multiple continuous views as a single integrated place, and linking the current view with long-term spatial knowledge. At the neural level, the PPA and RSC facilitate a coherent perception of the world, with the PPA representing the specific local geometry of the space and the RSC integrating multiple snapshots of views using spatiotemporal continuity and long-term memory.

14.6 Conclusions

Perceiving the geometry of space in our three-dimensional world is essential for navigating and interacting with objects. In this chapter, we have offered a review of key work in the behavioral, computational, and cognitive-neuroscience domains that has formalized space as an entity on its own. Space itself can be considered an "object of study," whose fundamental structure is composed of structural and semantic properties. We have shown that perception of the shape of a space is modulated by low-level image cues, top-down influences, stored knowledge, and spatial and temporal history. Like an object, a space has a function, a purpose, a typical view, and a geometrical shape. The shape of a space is an entity that, like the shape of an object or a face, can be described by its contours and surface properties. Furthermore, the perception of space is sensitive to task constraints and experience and subject to visual illusions and distortions in short-term and long-term memory. Lastly, evidence suggests that dedicated neural substrates encode the shape of space. Although the notion of studying space's "shape" may seem unorthodox, consider that, as moving agents, what we learn about the world occurs within a structured geometric volume of space.

Acknowledgments

The authors wish to thank Barbara Hidalgo-Sotelo and Madison Capps for insightful comments on the manuscript. This work was funded by National Science Foundation grants 0705677 and 0546262 to A. O. The author T. K. was funded by a National Science Foundation Graduate Research fellowship. All of the authors contributed equally to this chapter.

References

Aguirre G. K. and D'Esposito M. (1999). Topographical disorientation: a synthesis and taxonomy. *Brain*, 122: 1613–1628.

Bar, M. (2004). Visual objects in context. *Nature Rev. Neurosci.*, 5: 617–629.

Barnard, K. and Forsyth, D. A. (2001). Learning the semantics of words and pictures. In *Proceedings of the International Conference on Computer Vision*, Vancouver, Vol. II, pp. 408–415.

Bartlett, F. C. (1932). *Remembering: A Study in Experimental and Social Psychology*. Cambridge: Cambridge University Press.

Benedikt, M. L. (1979). To take hold of space: isovists and isovist fields. *Environ. Plan. B*, 6: 47–65.

Benedikt, M. L. and Burnham, C. A. (1985). Perceiving architectural space: from optic arrays to isovists. In W. H. Warren and R. E. Shaw (eds.) *Persistence and Change*, pp. 103–114. Hillsdale, NJ: Lawrence Erlbaum.

Bertamini, M., Jones L., Spooner, A., and Hecht, H. (2005). Boundary extension: the role of magnification, object size, context and binocular information. *J. Exp. Psychol.: Hum. Percept. Perf.*, 31: 1288–1307.

Bhalla, M. and Proffitt, D. R. (1999). Visual-motor recalibration in geographical slant perception. *J. Exp. Psychol.: Hum. Percept. Perf.*, 25: 1076–1096.

Burgess, N. (2006). Spatial memory: how egocentric and allocentric combine. *Trends Cogn. Sci.*, 10: 551–557.

Burgess, N., Becker, S., King, J. A., and O'Keefe, J. (2001). Memory for events and their spatial context: models and experiments. *Phil. Trans. R. Soc. Lond. B*, 356: 1493–1503.

Byrne, P., Becker, S., and Burgess, N. (2007). Remembering the past and imagining the future: a neural model of spatial memory and imagery. *Psychol. Rev.*, 114: 340–375.

Candel, I., Merckelbach, H., and Zandbergen, M. (2003). Boundary distortions for neutral and emotional pictures. *Psychonom. Bull. Rev.*, 10: 691–695.

Candel, I., Merckelbach, H., Houben, K., and Vandyck, I. (2004). How children remember neutral and emotional pictures: boundary extension in children's scene memories. *Am. J. Psychol.*, 117: 249–257.

Carson, C., Belongie, S., Greenspan, H., and Malik, J. (2002). Blobworld: image segmentation using expectation-maximization and its application to image querying. *IEEE Trans. Pat. Anal. Machine Intell.*, 24: 1026–1038

Chun, M. M. (2000). Contextual cueing of visual attention. *Trends Cogn. Sci.*, 4: 170–178.

Creem-Regehr, S. H., Gooch, A. A., Sahm, C. S. and Thompson, W. B. (2004). Perceiving virtual geographical slant: action influences perception. *J. Exp. Psychol.: Hum. Percept. Perf.*, 30: 811–821.

Cutting, J. E. (2003). Reconceiving perceptual space. In H. Hecht, M. Atherton, and R. Schwartz (eds.), *Perceiving Pictures: An Interdisciplinary Approach to Pictorial Space*, pp. 215–238. Boston, MA: MIT Press.

Cutting, J. E. and Vishton, P. M. (1995). Perceiving layout and knowing distances: the integration, relative potency, and contextual use of different information about depth. In W. Epstein and S. Rogers (eds.), *Perception of Space and Motion*, pp. 69–117, Vol. 5 of *Handbook of Perception and Cognition*. San Diego, CA: Academic Press.

Daniels, K. K. and Intraub, H. (2006). The shape of a view: are rectilinear views necessary to elicit boundary extension? *Vis. Cogn.*, 14: 129–149.

Davis, L. S. and Benedikt, M. L. (1979). *Computational Model of Space: Isovists and Isovists Fields*. Technical Report, School of Architecture, the University of Texas at Austin.

Epstein, R. A. (2008). Parahippocampal and retrosplenial contributions to human spatial navigation. *Trends Cogn. Sci.*, 12: 388–396.

Epstein, R. and Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature*, 392: 598–601.

Epstein, R., Graham, K. S., and Downing, P. E. (2003). Viewpoint-specific scene representations in human parahippocampal cortex. *Neuron*, 37: 865–876.

Epstein, R. A., Parker, W. E., and Feiler, A. M. (2007). Where am I now? Distinct roles for parahippocampal and retrosplenial cortices in place recognition. *J. Neurosci.*, 27: 6141–6149.

Ericson, J. and Warren, W. (2009). Rips and folds in virtual space: ordinal violations in human spatial knowledge. [Abstract.] *J. Vis.*, 9(8): 1143.

Fortenbaugh, F. C., Hicks, J. C., Hao, L., and Turano, K. A. (2007). Losing sight of the bigger picture: peripheral field loss compresses representations of space. *Vis. Res.*, 47:, 2506–2520.

Friedman, A. (1979). Framing pictures: the role of knowledge in automatized encoding and memory for gist. *J. Exp. Psychol.: Gen.*, 108: 316–355.

Gallagher, K., Balas, B., Matheny, J. and Sinha, P. (2005) The effects of scene category and content on boundary extension. In B. Bara, L. Barsalou, and M. Bucciarelli (eds.), *Proceedings of the 27th Annual Meeting of the Cognitive Science Society*. Stresa, Italy: Cognitive Science Society.

Greene, M. R. and Oliva, A. (2010). High-level aftereffects to global scene property. *J. Exp. Psychol.: Hum. Percept. Perf.*, 36(6): 1430–1442.

Grill-Spector, K., Henson, R., and Martin, A. (2005). Repetition and the brain: neural models of stimulus-specific effects. *Trends Cogn. Sci.*, 10: 14–23.

Held, R. and Banks, M. (2008). Perceived size is affected by blur and accommodation. [Abstract.] *J. Vis.*, 8(6): 442.

Hemmer, P. and Steyvers, M. (2009). A Bayesian account of reconstructive memory. *Topics Cogn. Sci.*, 1: 189–202.

Higashiyama, A. (1996). Horizontal and vertical distance perception: the discorded-orientation theory. *Percept. Psychophys.*, 58: 259–270.

Hochberg, J. (1978). *Perception*, (2nd edn). New York: Prentice-Hall.

Hochberg, J. (1986). Representation of motion and space in video and cinematic displays. In K. J. Boff, L. Kaufman and J. P. Thomas (eds.) *Handbook of Perception and Human Performance*, Vol. 1, pp. 22:1–22:64, New York: Wiley.

Hollingworth, A. and Henderson, J. M. (2004). Sustained change blindness to incremental scene rotation: a dissociation between explicit change detection and visual memory. *Percept. Psychophys.*, 66: 800–807.

Huttenlocher, J., Hedges, L. V., and Vevea, J. L. (2000). Why do categories affect stimulus judgment? *J. Exp. Psychol. Gen.*, 129: 220–241.

Indow, T. (1991). A critical review of Luneburg's model with regard to global structure of visual space. *Psychol. Rev.*, 98: 430–453.

Intraub, H. (2004). Anticipatory spatial representation in a deaf and blind observer. *Cognition*, 94: 19–37.

Intraub, H. and Dickinson, C. A. (2008). False memory 1/20th of a second later: what the early onset of boundary extension reveals about perception. *Psychol. Sci.*, 19: 1007–1014.

Intraub, H. and Richardson, M. (1989). Wide-angle memories of close-up scenes. *J. Exp. Psychol.: Learn., Mem. Cogn.*, 15: 179–187.

Intraub, H., Bender, R. S., and Mangels, J. A. (1992). Looking at pictures but remembering scenes. *J. Exp. Psychol.: Learn., Mem. Cogn.*, 18: 180–191.

Intraub, H., Hoffman, J. E., Wetherhold, C. J., and Stoehs, S. (2006). More than meets the eye: the effect of planned fixations on scene representation. *Percept. Psychophys.*, 5: 759–769.

Irwin, D. E., Zacks, J. L., and Brown, J. S. (1990). Visual memory and the perception of a stable visual environment. *Percept. Psychophys.*, 471: 35–46.

Kanizsa, G. and Gerbino, W. (1982). Amodal completion: Seeing or thinking? In J. Beck (ed.) *Organization and Representation in Perception*, pp. 167–190. Hillsdale, NJ: Erlbaum.

Koenderink, J. J., van Doorn, A. J., and Todd, J. T. (2009). Wide distribution of external local sign in the normal population. *Psychol. Res.*, 73: 14–22.

Konkle, T. and Oliva, A. (2007a). Normative representation of objects and scenes: evidence from predictable biases in visual perception and memory. [Abstract.] *J. Vis.*, 7(9): 1049.

Konkle, T. and Oliva, A. (2007b). Normative representation of objects: evidence for an ecological bias in perception and memory. In D. S. McNamara and J. G. Trafton (eds.), *Proceedings of the 29th Annual Cognitive Science Society*, pp. 407–413). Austin, TX: Cognitive Science Society.

Lappin, J. S., Shelton, A. L. and Rieser, J. J. (2006). Environmental context influences visually perceived distance. *Percept. Psychophys.*, 68: 571–581.

Leopold, D., O'Toole, A., Vetter, T., and Blanz, V. (2001). Prototype-referenced shape encoding revealed by high-level aftereffects. *Nature Neurosci.*, 4: 89–94.

Loomis, J. and Philbeck, J. (1999). Is the anisotropy of perceived 3D shape invariant across scale? *Percept. Psychophys.*, 61: 397–402.

Loomis, J. M., Da Silva, A., Fujita, N., and Fukusima, S. S. (1992). Visual space perception and visual directed action. *J. Exp. Psychol.: Hum. Percept. Perf.*, 18: 906–921.

Loomis, J. M., Da Silva, J. A., Philbeck, J. W. and Fukushima, S. S. (1996). Visual perception of location and distance. *Curr. Dir. Psychol. Sci.*, 5: 72–77.

Lyle, K. B. and Johnson, M. K. (2006). Importing perceived features into false memories. *Memory*, 14: 197–213.

Maguire, E. A. (2001). The retrosplenial contribution to human navigation: a review of lesion and neuroimaging findings. *Scand. J. Psychol.*, 42: 225–238.

Marr, D. (1982). *Vision*. San Francisco: W. H. Freeman.

McNamara, T. P., Rump, B., and Werner, S. (2003). Egocentric and geocentric frames of reference in memory of large-scale space. *Psychonom. Bull. Rev.*, 10: 589–595.

Michel, L. (1996). *Light the Shape of Space. Designing with Space and Light.* New York: Van Nostrand Reinhold.

Miller, M. B. and Gazzaniga, M. S. (1998). Creating false memories for visual scenes. *Neuropsychologia*, 46: 513–520.

Murray, S. O., Boyaci, H. and Kersten, D. (2006). The representation of perceived angular size in human primary visual cortex. *Nature Neurosci.*, 9: 429–434.

O'Keefe, J. A. (1979). A review of hippocampal place cells. *Prog. Neurobiol.*, 13: 419–439.

Oliva, A. (2005). Gist of the scene. In L. Itti, G. Rees, and J. K. Tsotsos (eds.), *The Encyclopedia of Neurobiology of Attention*, pp. 251–256. San Diego, CA: Elsevier.

Oliva, A. and Torralba, A. (2001). Modeling the shape of the scene: a holistic representation of the spatial envelope. *Int. J. Comp. Vis.*, 42: 145–175.

Oliva, A. and Torralba, A. (2002). Scene-centered description from spatial envelope properties. In H. H. Bülthoff, S.-W. Lee, T. Poggio, and C. Wallraven (eds.), *Biologically Motivated Computer Vision*, pp. 263–272. Lecture Notes in Computer Science, 2525. Berlin: Springer.

Oliva, A. and Torralba, A. (2006). Building the gist of a scene: the role of global image features in recognition. *Prog. Brain Res.: Vis. Percept.*, 155: 23–36.

Oliva, A., and Torralba, A. (2007). The role of context in object recognition. *Trends Cogn. Sci.*, 11: 520–527.

Oliva, A., Arsenio, H. C., and Wolfe, J. M. (2004). Panoramic search: the interaction of memory and vision in search through a familiar scene. *J. Exp. Psychol.: Hum. Percept. Perf.*, 30: 1132–1146.

Palmer, S. E. (1975). The effects of contextual scenes on the identification of objects. *Mem. Cogn.* 3: 519–526.

Park, S. and Chun, M. M. (2009). Different roles of the parahippocampal place area (PPA) and retrosplenial cortex (RSC) in panoramic scene perception. *Neuroimage*, 47: 1747–1756.

Park, S., Intraub, H., Widders, D., Yi, D. J., and Chun, M. M. (2007). Beyond the edges of a view: boundary extension in human scene-selective visual cortex. *Neuron*, 54: 335–342.

Park, S., Brady, T. F., Greene, M. R., and Oliva, A. (2011). Disentangling scene content from spatial boundary: complementary roles for the parahippocampal place area and lateral occipital complex in representing real-world scenes. *J. Neurosci.*, 31(4): 1333–1340.

Pirenne, M. H. (1970). *Optics, Painting and Photography*. Cambridge: Cambridge University Press.

Potter, M. C. (1976). Short-term conceptual memory for pictures. *J. Exp. Psychol.: Hum. Learn. Mem.*, 2: 509–522.

Proffitt, D. R., Bhalla, M., Gossweiler, R., and Midgett, J. (1995). Perceiving geographical slant. *Psychonom. Bull. Rev.*, 2: 409–428.

Proffitt, D. R., Stefanucci, J., Banton, T., and Epstein, W. (2003). The role of effort in perceiving distance. *Psychol. Sci.*, 14: 106–112.

Psarra, S. and Grajewski, T. (2001). Describing shape and shape complexity using local properties. In J. Peponis, J. Wineman and S. Bafna (eds.), *Proc. 3rd International Space Syntax Symposium*, pp. 28.1–28.16. Ann Arbor, MI: Alfred Taubman College of Architecture and Urban Planning, University of Michigan.

Ross, M. G. and Oliva, A. (2010). Estimating perception of scene layout properties from global image features. *J. Vis.* 10(1): 2.

Rothman, D. B. and Warren, W. H. (2006). Wormholes in virtual reality and the geometry of cognitive maps. [Abstract.] *J. Vis.*, 6(6): 143.

Schnall, S., Harber, K. D., Stefanucci, J. K., and Proffitt, D. R. (2008). Social support and the perception of geographical slant. *J. Exp. Social Psychol.*, 44: 1246–1255.

Schnapp, B. and Warren, W. (2007). Wormholes in virtual reality: what spatial knowledge is learned for navigation? [Abstract.] *J. Vis.*, 7(9): 758.

Seamon, J. G., Schlegel, S. E., Hiester, P. M., Landau, S. M., and Blumenthal, B. F. (2002). Misremembering pictured objects: people of all ages demonstrate the boundary extension illusion. *Am. J. Psychol.*, 115: 151–167.

Stamps, A. E. (2005). Enclosure and safety in urbanscapes. *Environ. Behav.*, 37: 102–133.

Tandy, C. R. V. (1967). The isovist method of landscape survey. In H. C. Murray (ed.) *Symposium: Methods of Landscape Analysis*, pp. 9–10. London: Landscape Research Group.

Torralba, A. and Oliva, A. (2002). Depth estimation from image structure. *IEEE Pattern Anal. Machine Intell.*, 24: 1226–1238.

Torralba, A. and Oliva, A. (2003). Statistics of natural images categories. *Network: Comput. Neural Syst.*, 14: 391–412.

Turner, A., Doxa, M., O'Sullivan, D., and Penn, A. (2001). From isovists to visibility graphs: a methodology for the analysis of architectural space. *Environ. Plan. B: Plan. Des.*, 28: 103–121.

Twedt, E., Hawkins, C. B., and Proffitt, D. (2009). Perspective-taking changes perceived spatial layout. [Abstract.] *J. Vis.*, 9(8): 74.

Vann, S. D., Aggleton, J. P., and Maguire E. A. (2009). What does the retrosplenial cortex do? *Nature Rev. Neurosci.*, 10: 792–802.

Wade, N. J. and Verstraten, F. A. J. (2005). Accommodating the past: a selective history of adaptation. In C. Clifford and G. Rhodes (eds.), *Fitting the Mind to the World: Adaptation and Aftereffects in High-Level Vision*. Advances in Visual Cognition Series?. Oxford: Oxford University Press.

Wagner, M. (1985). The metric of visual space. *Percept. Psychophys.*, 38: 483–495.

Watt, S. J., Akeley, K., Ernst, M. O., and Banks, M. S. (2005). Focus cues affect perceived depth. *J. Vis.*, 5: 834–862.

Webster, M. A., Kaping, D., Mizokami, Y., and Duhamel, P. (2004). Adaptation to natural facial categories. Nature, 428: 557–561.

Wiener, J. M. and Franz, G. (2005). Isovists as a means to predict spatial experience and behavior. In C. Freksa, M. Knauff, B. Krieg-Brückner, B. Nebel, and T. Barkowsky (eds.), *International Conference on Spatial Cognition 2004*, pp. 42–57. Lecture Notes in Computer Science, 3343. Berlin: Springer.

Witt, J. K., Proffitt, D. R. and Epstein, W. (2004). Perceiving distance: a role of effort and intent. *Perception*, 33, 570–590.

Woods, A. J., Philbeck, J. W., and Danoff, J. V. (2009). The various perceptions of distance: an alternative view of how effort affects distance judgments. *J. Exp. Psychol.: Hum. Percept. Perf.*, 35: 1104–1117.

Wu, B., Ooi, T. L., and Zijiang, J. H. (2004). Perceiving distance accurately by a directional process of integrating ground information. *Nature*, 428: 73–77.

Yang, T. L., Dixon, M. W., and Proffitt, D. R. (1999). Seeing big things: overestimates of heights is greater for real objects than for objects in pictures. *Perception*, 28: 445–467.