6 Representing, perceiving and remembering the shape of visual space

A. Oliva, S. Park and T. Konkle

Abstract

Humans constantly engage in automatic and rapid analysis of spatial scene structure when navigating an environment or searching for objects. This chapter draws on current research from computational, behavioral and neuroscience perspectives aimed at understanding how the human brain perceives, represents, and remembers the shape of space. Space can be described with both structural descriptions, which reflect layout of surfaces in the physical world, and semantic descriptions, which incorporate an observers understanding of the environment. We first review two formal approaches which can quantify the structural properties of scene layout: the isovist representation and the spatial envelope representation. Next we explore how space is experienced by observers, by reviewing behavioral results in which different factors distort the perceived space away from a veridical representation. Finally, we examine how representations of specific views of space are maintained in memory, and discuss potential neural mechanisms involved in integrating views into larger environments. Gaining a formal understanding of how geometric aspects of space map onto human perceptual, cognitive, and neural systems will help to create efficient functional spaces for our daily interactions with the world.

6.1 Introduction

Recognizing the current environment determines our ability to act strategically, for example when selecting a route for walking, anticipating where objects are likely to

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

appear, and knowing what behaviors are appropriate in a particular context.

Whereas objects are typically entities we act upon, environments are entities 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, Zacks, and Brown, 1990; Oliva, Arsenio and Wolfe, 2004; Park and Chun, 2009), and with what has been stored over the 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 a space in a three dimensional physical world, but we often study our perception of space through two-dimensional pictures. While there are likely important differences between the perception of space in the world and the perception of space mediated through pictures, we choose to describe in this chapter principles that are likely to apply to both mediums. In the following sections, we begin by describing how properties of space can be formalized, and to which 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 peoples memory of scenes and 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.

Visual space perception 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 the cone of visible space as seen from an observers vantage: a view is oriented (e.g. looking up, straight, down) and has an aperture that the dioptrics of the eyes suggest covers up to 180 degrees. However, the apparent visual field that human observers visually experience is closer to 90 degrees, corresponding to a hemisphere of space in front of them (Koenderink, van Doorn and Todd, 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 the observer is looking at or embedded in (semantic level).

While in physics, space is defined as the opposite of mass, in our structural level description, we define a space as an entity composed by two substances: mass and holes. A space can be of any physical size in the world, i.e. 1 m3 or 1000 m3. 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. From that location, if the observer rotates through 360 degrees, the set of visible surfaces visible from those viewpoints

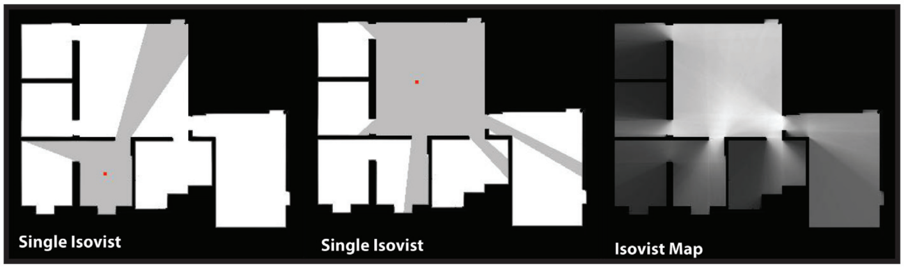

Figure 6.1. Two single isovist views are shown, with the red dot marking the location from which the isovist was generated. The right most image shows the isovist map, which is the collection of isovists from all possible locations within the space.

is called an isovist (Benedikt, 1979). An example of an isovist is shown in Figure 6.1. A collection of all isovists 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 with a statistical representation of the image featuresthis statistical representation can coarsely describe 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 the specific physical space (e.g. my kitchen, the White House) or groups of physical spaces (e.g. the category of industrial kitchens or gymnasiums).[1] The term "scene" has two common usages in the literature as both a particular view as well as 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 ones view of a space (unlike places, which do not depend on the viewpoint of the observer). Therefore, a place can be composed by 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 views. Places and scenes can be conceptualized as part of a larger topology, an environment. Environments would therefore typically refer to physical spaces encompassing one or more scenes and places, typically of a larger scale than a single place.

[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* that are hippocampus neurons which fire when an animal is at a particular three-dimensional location (e.g., O'Keefe, 1070). Place-cells are specified by latitude, longitude, and height coordinates, and can also be oriented, e.g., pointing north with a 30 degree downward angle.

6.2 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 3-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 texture of surfaces.

6.2.1 Isovist representation

Figure 6.1 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 shape of the place, at a given location, obtained from the observer rotating through 360 degrees.

A concept initially introduced by Tandy (1967), *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 6.1. The volumetric configuration of a place requires calculating a collection of isovists at various locations: this refers to *isovist field* or *isovist map* (Benedikt, 1979; Davis and Benedikt, 1979), shown in Figure 6.1 (right). High luminance levels indicate areas that can be seen from the most locations of the main central room of the lab, and dark areas indicate regions that are hidden from most of the locations. In empty and convex rooms (like a circle, squared or rectangular room), the isovist field is homogeneous, as every isovist from each location has the same shape and volume (or has the same area if considering a two-dimensional floor plan).

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 the 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 such 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 intersection of the outlines of the isovist polygon, or the openness of the polygon. Isovist openness is calculated as the ratio between open edges (generated by occlusions) and closed edges (defined by solid visible boundaries, Psarra and Gradjewski, 2001; Wiener and Franz, 2005).

From simple geometrical measurements of isovist and isovist maps, higher level

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

properties of the space can be derived: its occlusivity (i.e. the depth to which surfaces inside the space are overlapping with each other, Benedikt, 1979[3]), its degree of compactness (a measure defined by a circle whose radius is equivalent to the isovist mean radial length which gives an account of how much the isovist's shape resembles a circle), its degree of spaciousness (Stamps, 2005), and its degree of convexity (also referred as jaggedness, calculated as the ratio between isovist squared perimeter and its area, see Wiener and Franz, 2005; Turner et al., 2001). A concave or "jagged" isovist would have dents, which would mean that regions of the place are hidden from view. A circular and convex isovist would have no hiding regions.

Our understanding of the relationship between geometrical measurements of the isovists and the perception of a scene and place remains in its infancy. Wiener and Franz (2005) found that the degree of convexity and openness ratio of isovists correlate 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), as well as 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 the function of the space. Further, behavior in a space may be predicted by these structural spatial descriptors. Along these lines, Turner and colleagues were able to predict complex social behaviors such as way-finding and movement of a crowd in a complex environment (Turner, et al., 2001).

An analysis of the kinds of space a human being encounters, and the geometrical properties that distinguish between the different kinds of spaces, remain to be done. Further, in its original form, the isovist theory does not account for the type of textures, materials, or colors attached to the surfaces, and this information will likely be important when relating structural descriptions of space to human perceptions of space 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 or 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.

6.2.2 Spatial envelope representation

Given that we experience a three-dimensional world, it makes sense that we 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, its 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 the whole space that provides an "instant impression of the volume of a room or an urban site" (Michel,

[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.

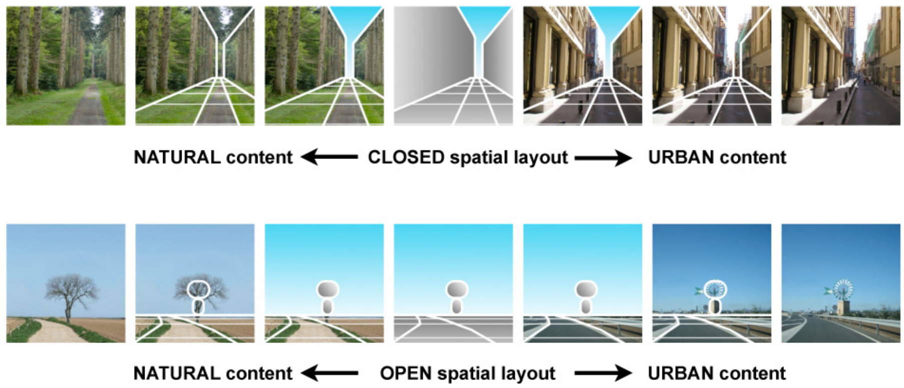

Figure 6.2. A schematic illustration of how pictures of real world scenes can be defined uniquely by their spatial layout and content. Note that the configuration, size and locations of components can be in correspondence between natural and manufactured environments. Keeping the enclosed spatial layout, if we strip off the natural content of a forest and fill the space with urban contents, then the scene becomes an urban street scene. Keeping the open spatial layout, if we strip off the natural content of a field and fill the space with urban contents, then the scene becomes an urban highway.

1996). The concept has been used to describe qualitatively the character and mood of a physical or a pictorial space, represented by its boundaries (e.g., walls, floor, ceiling, and lighting) stripped of movable elements (e.g., objects, furnishing).

In 2001, Oliva and Torralba extended the concept and proposed a formal, computational approach to the capture of the shape of space, as it would be perceived from an observers vantage (Oliva and Torralba, 2001, 2002, 2006, 2007; Torralba and Oliva, 2002, 2003). The collection of properties describing the space in view refers to 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 contents.

To give an example of these scene properties, a space can be represented by two independent descriptors, one representing the boundaries or external features, and one representing the content or internal features (Oliva and Torralba, 2001, 2002, 2006; Park et al., submitted). 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 6.2 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, 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, spatial resolution) diagnostic of a particular spatial layout (Oliva and Torralba 2001; Torralba and Oliva, 2002; Ross and Oliva, in press). 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 6.3 below the original image, where random noise was coerced to have 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 6.4. For simplicity, the model is presented here as a combination of four global scene properties (Figure 6.4A). The implementation of the model takes the form of high-level image filters originating from the outputs of local oriented filters, as in early visual areas of the brain (Figure 6.4B). 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 perspective) and the properties of its content (e.g. the style of the surface, natural or man-made, 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 along 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 describing a forest as an environment with trees, bushes and leaves, a forest 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 a "man-man outdoor place with perspective and medium level of 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 scenes images judged by people to have the same categorical membership (street, highway, forest, coastline, etc.) were projected close together in a multi-dimensional space whose axes correspond to the Spatial Envelope dimensions (Figure 6.4C). Neighborhood images in the spatial envelope space corresponded to images with similar spatial layout as well as a similar semantic description (more so when the space is filled densely, i.e. with either a lot of varied exemplars, or typical exemplars of categories).

As shown in these sections, both the Isovists and 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 surfaces, like 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).

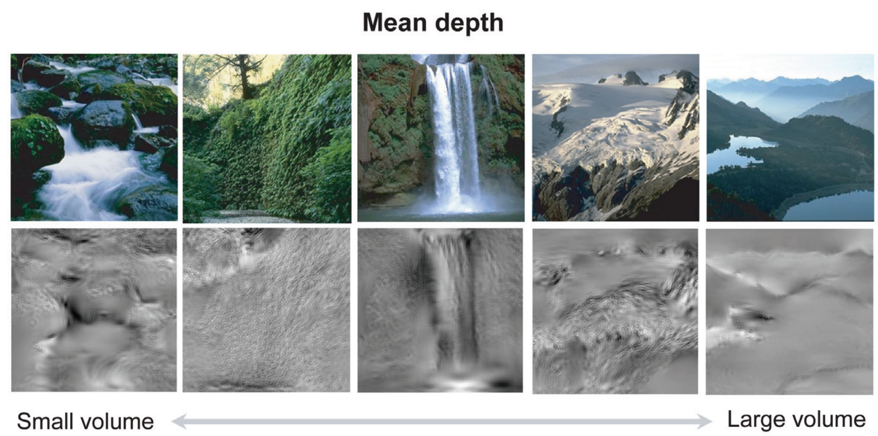

Figure 6.3. Top: Examples of natural scene images with different degree of mean depth (from small to large volume). Bottom: a sketch representation of the visual features captured with the spatial envelope representation (see Oliva and Torralba, 2001, 2006 for details). Note that this representation of a natural scene has no explicit coding of objects or segmented regions.

6.3 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 move 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, distance or slants of a space, rather than local perception centered on objects or parts.

6.3.1 Distortion of space geometry

There are many ways in which our perception of space is not veridicale.g. distances in the frontal plane (i.e. traversing from left to right) appear much larger than distances in the sagital 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; Yank, Dixon, and Proffitt, 1999). Surface angles are often underestimated and slants of hills are often overestimated (Proffitte et al., 1995; Creem-Regehr et al., 2004). Distances to objects can be misperceived when a relatively wide expense of the ground surface is not visible (Wu, Ooi, and Zijiang, 2004), or when the field of view is too narrow (Fortenbaugh, et al., 2007). Such biases

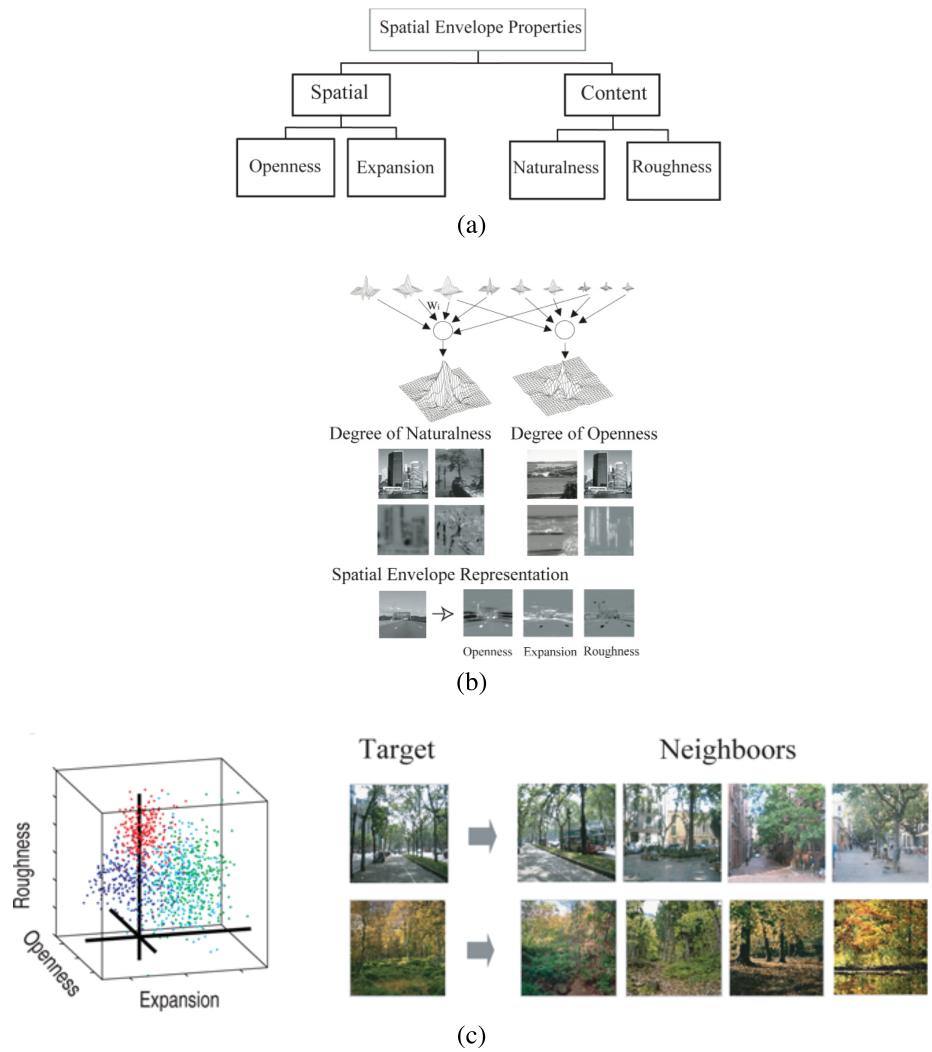

Figure 6.4. Schematic Spatial Envelope model. (a) Spatial envelope properties can be classified into spatial and content properties. (b) Illustration of a computational neuroscience implementation of the Spatial Envelope model. Features of naturalness and openness are illustrated here. (c) Projection pictures of man-made 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 targets images with their nearest neighbors in the Spatial Envelope space are shown here (from a dense database of images).

are also highly dependent of the scene structure: distance judgments are most difficult and inaccurate in a corridor, are easy, accurate, and reliable out on an open field, and are easy and reliable but inaccurate in a lobby (Lappin, Shelton, and Rieser, 2006). Many visual illusions, e.g. 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 distorting the perception of physical space have been well documented (for a review, Cutting, 2003): as physical distance increases, perceived distances are foreshortened as compared to physical space (Loomis and Philbeck, 1999). This means that observers do not accurately evaluate distance between objects at far distance, being only able to judge ordinal relations (which surface is in front of another, but not by how much). The compression of space planes with distance of viewing is likely due to the decrease in available information and depth cues (Indow, 1991, see Cutting, 2003). In his 2003s review, Cutting reports three classes of ecological space perception ranges. First, perception in the personal space (up to about two to three meters), is metric: indeed, veridical space 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 practically defined by the distance to which one can throw an object accurately (up to about 30 meters away). Whereas depth perception in action space suffers some compression, studies found it to be close to physical space. Beyond a few tens of meters is vista space, where observers perception of distances to and between surfaces can become greatly inaccurate, with a dramatic accelerating foreshortening of space perception with distances over 100 meters. 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 knowledge of relative size between objects, and ordinal cues like layout arrangement and occlusion to infer the shape of the three dimensional space. Further, these drastic spatial compressions of vista space are not noticed by individuals (Cutting and Vishton, 1995; Cutting, 2003).

This gradient of perceived space compression suggests the need for a perceptual isovist, where the characteristics of the shape of visible surfaces is measured not from actual distances in a volume but perceived distances (see Section 6.2.1). For example, we would expect a more deformed isovist for background than foreground planes, where perception is based on ordinal estimations. When only ordinal depth planes information is available, some illusions of scene volume and misinterpretation of surfaces may occur. Figure 5 illustrates such illusions using photographs of natural scenes: the "mountain cliff" and the picture of the "river receding into the distance" (Figure 6.5, a and c), are perceived respectively as "the base on a mountain" and "a view looking up at the sky" (Figure 6.5, b and d) 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 it produces in some cases large changes in the perceived scale of the space. The spatial envelope approach (Oliva and Torralba, 2001; Torralba and Oliva, 2002) captures the low level and texture statistics which are correlated with the change of perceived scale and semantic.

(a) (b) (c) (d)

Figure 6.5. Examples of natural images in which inverting the images creates a plausible scene view with dramatic changes in interpretation of surfaces and volume between upright and upside down (adapted from Torralba and Oliva, 2003).

Figure 6.6. 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.

6.3.2 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 (see Figure 6). The degree of focus across a scene is a simple low-level depth cuee.g. 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 at an object very close in front of you, will lead to only a small portion of the image that can be in focus with 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 (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. Original special effects took advantage of the 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 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 as well as our expectations about the structure and statistics of the natural world. Indeed, neuroimaging work has shown that the size you think something is in the world matters beyond just the visual angle that it projects on the retina. Murray, Boyaci, and Kersten (2006), presented observers with two discs of matched visual angle on the screen, but with contextual information that made one disc look much larger (and farther away) than the other. The bigger disc activated a greater extent of primary visual cortex than the smaller disc did, despite their equivalent visual size. These results suggest the perceived physical size of an object or space has consequences on very early stages of visual processing.

6.3.3 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) show that non-optical cognitive variables may influence the perception of space cues.

Along with task constraints, physical resources and capabilities available to the agent change the perception of space (e.g., distance, 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 participants intended to walk the distance (Witt, Proffitt, and Epstein, 2004). Similarly, backpack load manipulation had no effects on people's distance estimation when they were asked to throw a ball in the direction of the object. However, the weights of a 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, Proffitt, and Epstein, 2004; but see also Woods et al., 2009).

Other studies have shown that inherent characteristics or physical capabilities of an individual can also influence how they perceive space. For example, compared to 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, Hawkins, and Proffitt, 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 steep than participants who imagined neutral and negative social contact (Schnall, et al., 2008). Although it is hard to conclude from these studies whether these non-optical factors fundamentally changed the observers' perceptions or whether they modulated responses without changing perception, they provide evidence that the experience of geographical properties of the space can be influenced by changes in psychological load of an observer, beyond attributes of the physical world.

6.3.4 Adaptation to spatial layout

The previous sections presented examples of how spatial low-level cues and preexisting top-down knowledge can influence the 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 spacethe experience you had with particular visual scenes just moments ago can change the 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, the adaptation of those features affects the conscious perception of a subsequently presented stimulus (e.g. 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 (in press), tested whether observers adapt to global spatial envelope properties (described in the spatial envelope section 6.2.2), such as mean depth and openness. In one study, observers were presented with a stream of natural scenes which were largely different (categories, colors, layout, etc.), but which were all exemplars of very open scenes, representing vista space (panorama views on field, coastline, desert, beach, mountain, etc). Following this adaptation phase, a scene picture with a medium level of openness (e.g. a landscape with 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. On the contrary, the same ambiguous test images were judged to be open following adaptation to a stream of closed scenes (e.g. like cave, forest, canyon). Similar aftereffects held after adapting to other spatial envelope properties poles such as small versus large depth, natural versus urban spaces, and even to higher level properties of the scene, such as if the view depicts an environment with a hot versus cold climate.

Importantly, Greene and Oliva further showed that adaptation to different scene envelope properties not only influence judgments of the corresponding scene properties on a new image, but also influence categorical judgments on a new image. This experiment took advantage of the fact that fields are usually open scenes, while forests are typically closed scenes, but importantly there is a continuum between field and forest scenes, with some scenes existing ambiguously between the two categories that can be perceived both as a field or a forest (see Figure 6.7).

During the adaptation phase, observers were presented with a stream of images which again varied in basic level category and surfaces features, but were all depicting open views or closed views. No forests or fields were presented in this stream of images. After observers have adapted to for instance, open scenes, an ambiguously open or closed image should appear to be more closed. The critical question was whether an ambiguous field/ forest image will also appear to be more likely judged as a forest than a field, which has a more enclosed property. Similarly, adapting to a stream of closed natural images should cause an ambiguous field/forest to be judged as

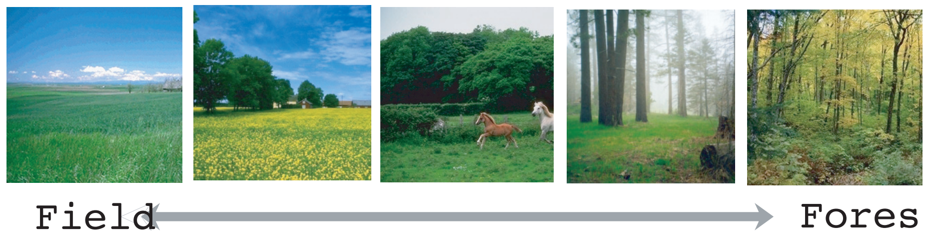

Figure 6.7. Continuum between forests and fields. Images in the middle of continuum have an ambiguous category and can be perceived as both a field and a forest.

more likely to be a field. Indeed, this is exactly what Greene and Oliva observed.

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

6.4 Remembering the shape of a space

The previous section reviewed evidence that the perception of space can be manipulated by low-level image cues, top-down influences, and temporal history with scenes. These perceptual illusions occur on-line while the relevant sensory information is present in the world, but similar systematic distortions of space occur as we represent scene information that is no longer in view but 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.

6.4.1 Behavioral and neural aspects of boundary extension

When presented with a scene view, what do observers remember about the depicted space? Intraub and Richardson (1989) presented observers with pictures of scenes, and found that when observers drew the scenes from memory, they systematically drew more of the space than was actually shownthis 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 different tasks beyond drawing, such as rating and border-adjustment (e.g. Intraub, Bender and Mangels, 1992; Intraub, et al., 2006), different image sets (Candel, Merckelbach, and Zandbergen, 2003; Daniels and Intraub, 2006), operates over a range of time scales

(a)

(b)

Figure 6.8. Example of Boundary Extension. After viewing a close-up view of a scene (a), observers tend to report an extended representation (b).

from minutes to hours (Intraub and Dickenson, 2008), and is found in young children as well as older adults (Candel et al., 2004; Seamon et al., 2002). Interestingly, boundary extension occurs even when observers are blindfolded they explore space with their handssuggesting an important link between the representations of space across sensory modalities (Intraub, 2004). Figure 6.8 shows an example of boundary extension. Observers presented with the scene in Figure 8a will remember the scene as having more information around the edges, depicted in Figure 6.8b.

In a functional neuroimaging study, Park and colleagues (Park, et al., 2007) examined whether scene selective neural regions showed evidence of representing the space encompassed within scene view. Critically they used a neural adaptation paradigm (also called repetition attenuation; Grill-Spector, Henson and Martin, 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 to when it was processed as a novel stimulus. This logic suggests that a second presentation of the stimulus matched what 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 the predictions of boundary extension. When presented with a close scene view, boundary extension predicts that this scene view might be represented at a wider angle than 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 those stimuli is reversed, the representation of the wide-angle view will be very different than 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, shown in Figure 6.9.

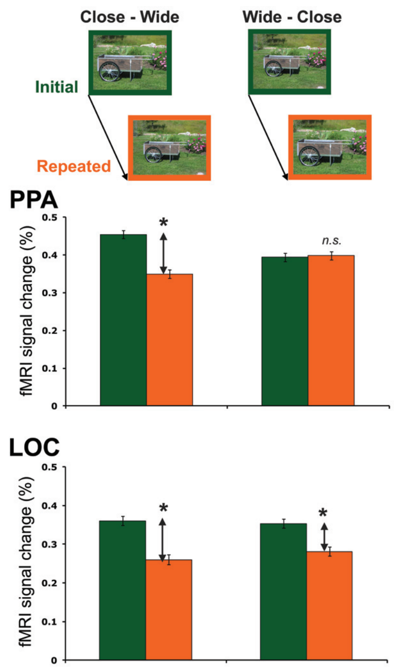

Figure 6.9. Examples of close-wide and wide-close condition are presented in the top row.Peaks of the hemodynamic responses for close-wide and wide-close conditions are shown for the PPA and LOC. An interaction between the close-wide and wide-close condition activation, representing boundary extension asymmetry, was observed in the PPA but not in the LOC. Error bars indicate standard error (± s. e. m.). Figure adapted from Park et al., 2007.

6.4.2 Navigating to remembered scene views

While boundary extension can be interpreted as an extrapolation of information in the scene periphery (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 memory of a view from a specific location might be influenced by the prototypical view.

In general, a view of a scene arises from an observers location in a three-dimensional space. As the observer walks through an environment, the view 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, 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, see Figure 6.10). However, there also are many views with the same scene gist (e.g. 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

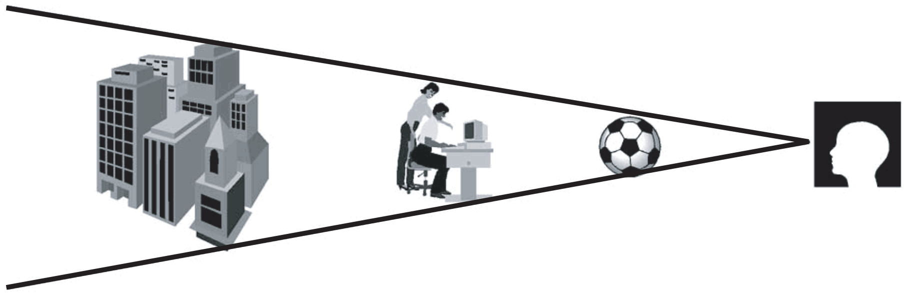

Figure 6.10. The semantic meaning of a scene changes as the depth of the scene increases. From a large distance, an observer may view buildings, which on the approach, change to rooms, or singleton objects on a surface. Figure adapted from Torralba and Oliva, 2002

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 store 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 scene semantic gist. Two places are shown in Figure 6.11, with the closest possible view (left), the farthest possible view (right), and the preferred view across observers (middle).

Despite the subjectivity of the task, 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 "e.g. 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 wanted to step either left or right, which was not allowed in the current experimental design. These data suggest that given a scene, some views are indeed better than others, and observers have a sense of how to walk to 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 from 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, observers were placed in the room at either the back or front of the space and had to maneuver through the space to match where they stand 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 op-

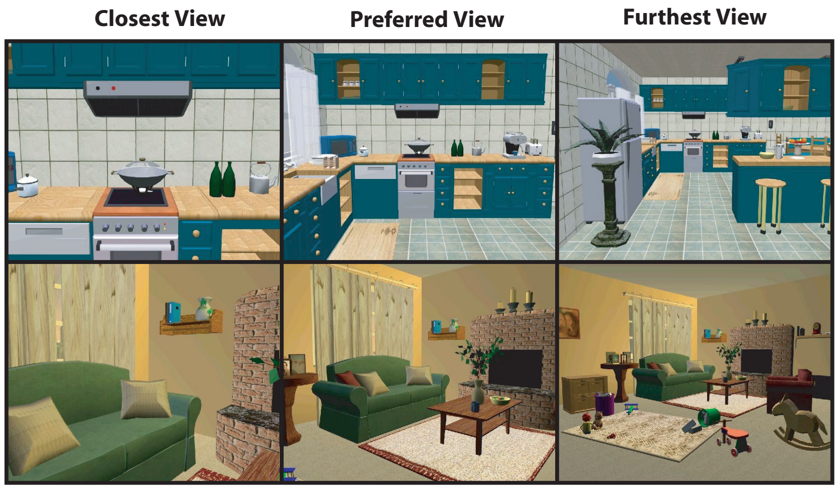

Figure 6.11. Two example spaces, a kitchen (top) and living room (bottom). The closest and farthest views of the scenes are shown (left and right, respectively), as well as the preferred scene view across observers (middle).

posite pattern was observed, where people tended to navigate to a closer location than the studied view (see Figure 6.12). Thus in this experimental task with these scenes, we observed that both boundary extension for close views and boundary restriction for far views. Importantly, memory errors were not 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 1 to 2 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 6.4.2).

These systematic biases in memory can be explained by the notion that memory for a scene is reconstructive (e.g. Bartlett, 1932). On this account, when an observer has to navigate to match the scene view in memory, if they have any uncertainty about the location, then they should not guess randomly from among those options, but instead should choose the view that is closer to the prototypical view. This way, memory for a particular scene representation can take advantage of the regularities observed over 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, Hedges, and Vevea, 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) as well as the complexity of the background scene (e.g. Gallagher et al., 2005). The relative contribution of structural space features and semantic scene features in these effects is unknown, and we cannot yet take an ar-

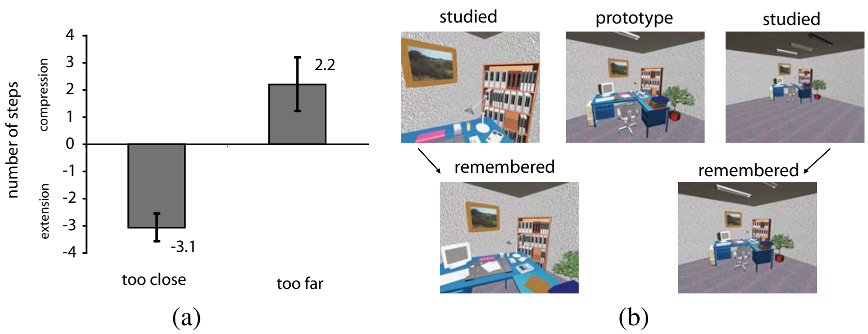

Figure 6.12. A) Memory errors for scenes presented too close or too far, measured in number of steps. Error bars reflect ±1 S. E. M. B) Example scene used in the memory experiment. Observers were presented with either a too close or too far view. The remembered scene views were systematically biased towards the prototypical view.

bitrary scene and understand what spatial distortions will be present in memory. We believe that such predictions will be possible as we gain a richer and quantitative vocabulary that characterizes the many possible spatial relationships between an observer and the elements in front of 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 degrees. However, in natural viewing conditions, observers navigate through the environment with a full field view. While previous studies have demonstrated that shape of the aperture (rectilinear, oval, irregular) does not effect 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 in a full-field display (see Figure 6.13; Konkle, Boucart, and Oliva, unpublished data). We found that 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.

The data from the panoramic study demonstrate that these scene memory mechanisms discovered from 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 memory of a specific view, and that scene processing mechanisms may not only serve to help construct a continuous world, but also to support optimal views for perception and memory of a three dimensional space.

6.5 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

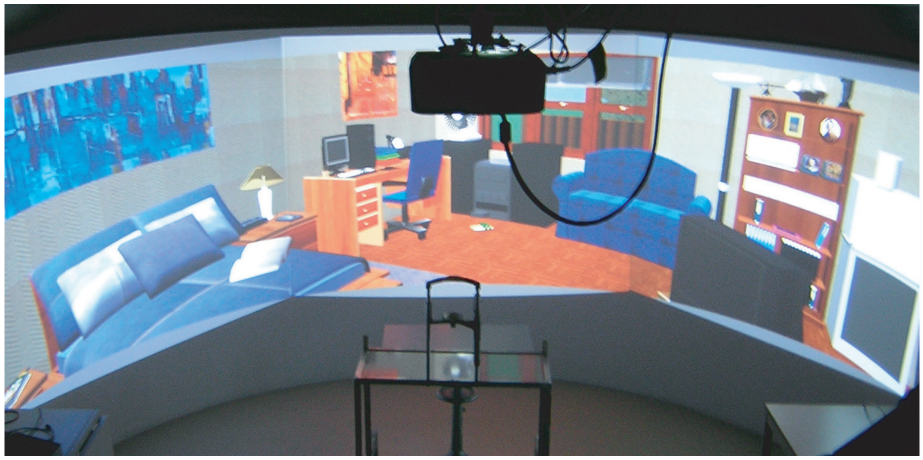

Figure 6.13. Scene view presented in a full field display. Observers were seated at the table with their head position fixed by the chin rest.

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, Zacks, and Brown, 1990; Lyle and Johnson, 2006; Miller and Gazzaniga, 1998, Oliva, Arsenio and Wolfe, 2004), and by extrapolating scene boundaries (Intraub and Richardson, 1989; Hochberg, 1986, see Section 6.5). 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 the place beyond the single scene in view. In the fMRI scanner, participants were presented with three consecutive and overlapping views from a single panoramic scene (see Figure 6.14), so that the observers perceived a natural scan of the environment, moving their head from left to right. Researchers investigated whether brain regions known to respond preferentially to pictures of natural scenes and spaces also show sensitivity to views that were 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, Brady, Greene and Oliva, submitted), has a view-specific representation (see also Epstein, Graham, and Downing, 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, Aggleton, and Maguire, 2009 for 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 to view 3. Additional experiments suggested that the RSC only showed such neural attenuation when the views were displayed in close spatio-temporal continuity. When the same

Figure 6.14. Panoramic 1st and Panoramic 3rd image were taken from a single panoramic view. Panoramic 1st, 2nd, and 3rd images were sequentially presented one at a time at fixation. The 1st and the 3rd image overlapped in 33% of its physical details.

trials were presented with a longer lag or intervening items between views, the RSC no longer showed the 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., scanning the world through simulated head and eye movements while the viewers location is constant), Epstein, Parker, and Feiler (2007) tested the neural basis of extrapolating views to a larger volume, beyond the viewer's current location. In their study, they presented participants from the University of Pennsylvania community with views of familiar places around the campus or views from a different, unfamiliar campus. Participants tasks were to judge the location of the view (e.g., whether on the west or east of 36th street), or its orientation (e.g., whether facing west or east of the campus). Whereas the PPA responded equally to all conditions, RSC activation was strongest to location judgments. This task required information about the viewers current location, as well as the location of the current scene within the larger environment. RSC activation was second highest when 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. 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 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. RSCs involvement in the retrieval of long term memory is consistent with patient and neuroimaging studies that have shown the involvement of RSC in episodic and autobiographic memory retrieval (Burgess et al., 2001; Maguire, 2001; Byrne, Becker, and Burgess, 2007).

In a related vein, several spatial navigation studies suggest that people can use geometric environmental cues such as landmarks (Burgess, 2006; McNamara, Rump and Werner, 2003), or alignments to the walls to recognize a novel view of the same place as fast as a learned view, suggesting they 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 geometric relationship between places within a given environment. For instance, subways acts like "wormholes" (Rothman and Warren, 2006; Schnapp and Warren, 2007; Ericson and Warren, 2009), distorting the perception of spatio-temporal relationships between 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 in between two objects at different locations within a maze, they naturally took advantage of wormhole shortcuts and avoided going around a longer path in a maze. Observers did not notice that wormholes violated the Euclidean structure of the geometrical map of the maze. These results demonstrate that the spatial knowledge of a broad environment does not exist as a complete integrated cognitive map per se, but exists as combination of local neighborhood directions and distances embedded in a weak topological structure of the world.

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

6.6 Conclusion

Perceiving the geometry of space in our three dimensional world is essential for navigating and interacting with objects. In this chapter, we offer a review of key work in behavioral, computational and cognitive neuroscience domains which have 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 show that perception of the shape of space is modulated by low level image cues, top down influences, stored knowledge as well as spatial and temporal history. Like an object, a space has a function, a purpose, a typical view, and a geometrical shape. The shape of space stands as an entity that, like the shape of an object or a face, can be described by its contours and surface properties. Furthermore, space perception 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 shape space. Although the notion of studying spaces "shape" may seem unorthodox, consider that, as moving agents, what we learn about the world occurs within the structured geometric volume of space.

Acknowledgements

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 0705677 and 0546262 to A. O. T. K. is funded by a National Science Foundation Graduate Research fellowship. The authors contributed equally to this chapter. Correspondence may be sent to any author: A.Ø. (oliva@mit.edu), S. P. (sjpark31@mit.edu) or T. K. (tkonkle@mit.edu).

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 Reviews Neurosci.*, 5: 617–629.

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

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

Benedikt, M. L. (1979). To take hold of space: isovists and isovist fields. *Environment and Planning 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, Lawrence Erlbaum: Hillsdale, NJ)

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 Cognit. Sci.*, 10: 551–557.

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.

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.

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

andel, 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 Intel.*, 24: 1026–1038

Chun, M. M. (2000). Contextual cueing of visual attention. *Trends Cognit. 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*, 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.) *Handbook of Perception and Cognition: Vol. 5. Perception of Space and Motion*, pp. 69–117. 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? *Visual 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. 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. Vision*, 9: 1143a.

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. *Vision 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. (in press). High-Level Aftereffects to Global Scene Property. *J. Exp. Psychol.: Hum. Percept. Perf.*

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.

Hemmer, P. and Steyvers, M. (2009). A Bayesian Account of Reconstructive Memory. *Topics Cognit. Sci.*, 1: 189–202.

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

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

Hochberg, J. (1978). *Perception (2nd edn)*, 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, 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. *Cognit.*, 94: 19–37.

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

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

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.

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.

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

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.

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.

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. M., Da Silva, A., Fujita, N. and Fukusima, S. S. (1992). Visual Space Perception and Visual Directed Action. *J. Exp. Psychol.: Hum. Percept. Perf.*, 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.

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

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

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

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, CA: W. H. Freeman.

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*, pg. 2510-256, San Diego, CA: Elsevier.

Oliva, A., Arsenio, H. C., 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.

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. *Lecture Notes In Computer Science*, 2525: 263–272.

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

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

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

Park, S., Brady, T. F., Greene, M. R. and Oliva, A. Disentangling scene content and scene layout: Complementary roles for the PPA and LOC in representing natural images. submitted.

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.

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 geographicalslant. *Psychonom. Bull. Rev.*, 2: 409–428.

Proffitt, D. R., Stefanucci, J., Banton, T. and Epstein, W. (2003). The role of effort in perceiving distance. *Psycholog. 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. Alfred Taubman College of Architecture and Urban Planning, University of Michigan.

Ross, M. G. and Oliva, A. (in press). Estimating perception of scene layout properties from global image features. *J. Vision*.

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

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. Vision*, 7: 758a.

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. *Environment and 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.

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

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

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

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

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. Volume 2, Advances in Visual Cognition Series*. Oxford: Oxford University Press.

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

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

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 Spatial Cognition 2004*, Vol. 3343 of Lecture Notes in Computer Science, pp. 42–57. Berlin, Germany: Springer.

Witt, J. K., Proffitt, D. R. and Epstein, W. (2004). Perceiving distance: A role of effort and intent. *Percept.*, 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.: Human 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. *Percept.*, 28: 445–467.