Intraub, H. (2014). Visual scene representation: A spatial-cognitive perspective. In K. Kveraga & M. Bar (Eds.), Scene vision: Making sense of what we see (pp. 5-26). Cambridge, MA: The MIT Press.

1

Visual Scene Representation: A Spatial-Cognitive Perspective

Helene Intraub

Traditionally, scene perception has been conceptualized within the modality-centric framework of visual cognition. However, in the world, observers are spatially embedded within the scenes they perceive. Scenes are sampled through eye movements but also through movements of the head and body, guided by expectations about surrounding space. In this chapter, I will address the idea that scene representation is, at its core, a spatio-centric representation that incorporates multiple sources of information: sensory input, but also several sources of top-down information. Boundary extension (false memory beyond the edges of a view; Intraub 2010; Intraub & Richardson, 1989) provides a novel window onto the nature of scene representation because the remembered "extended" region has no corresponding sensory correlate. I will discuss behavioral, neuroimaging and neuropsychological research on boundary extension that supports a spatio-centric alternative to the traditional description of scene representation as a *visual* representation. I will suggest that this alternative view bears a relation to theories about memory and future planning.

The traditional modality-centric approach to scene representation continues to generate interesting questions and valuable research, but may unnecessarily constrain the way we think about scene perception and memory. A key motivation underlying much of the research on visual scene perception has been the mismatch between the phenomenology of vision (perception of a coherent, continuous visual world) and the striking limitations on visual input imposed by the physiology of vision (O'Regan, 1992). Put simply, the world is continuous, but visual sensory input is not. The visual field is spatially limited. To perceive our surroundings we must sample the world through successive eye fixations and movements of the head and body. Ballistic eye movements (saccades) shift the eyes' position between fixations, and during these eye movements, vision is suppressed (Volkmann, 1986). Thus, the currently available information during scene perception switches between visual sensory perception and transsaccadic memory (Irwin, 1991) as frequently as three times per second. Finally, each time the eyes land, our best visual acuity is limited to the tiny foveal region

(1° of visual angle) of each eye and declines outward into the large low-acuity periphery of vision (Rayner, 2009; Rayner & Pollatsek, 1992). How this piecemeal, inhomogeneous input comes to support the experience of a coherent visual world has been one of the classic mysteries of visual scene perception (Hochberg, 1986; Intraub, 1997; Irwin, 1991).

Rather than thinking of this problem in terms of the visual modality alone, an alternative approach is to consider that visual scene perception, even in the case of a 2D photograph, may be organized within the observer's spatial framework of surrounding space (the multisource model: Intraub, 2010, 2012; Intraub & Dickinson, 2008). Here, the underlying framework for scene perception is the observer's sense of space (e.g., "in front of me," "to the sides," "above," "below," and "behind me"; Tversky, 2009). This spatial framework acts as scaffolding that organizes not only the visual input but also rapidly available sources of information about the likely world from which the view was taken. These other sources of information include amodal completion of objects (Kanizsa, 1979) and amodal continuation of surfaces (Fantoni, Hilger, Gerbino, & Kellman, 2008; Yin, Kellman, & Shipley, 2000) that are cropped by the boundaries of the photograph; knowledge based on rapid scene classification (occurring within 100–150 ms of stimulus onset; Greene & Oliva, 2009; Potter, 1976; Thorpe, Fize, & Marlot, 1996); as well as object-to-context associations (Bar, 2004). The ability to rapidly identify objects and scenes provides early access to expectations (and constraints) about the likely layout and content of the surrounding world that a single view only partially reveals.

During day-to-day interactions with the world, the observer is embedded within a surrounding scene (e.g., standing in a kitchen) with online access to one view at a time. Scene representation, in this conceptualization, captures this fundamental reality. A single view (e.g., the first view on a scene, or the frozen view presented in a photograph) is thought to activate multiple brain areas that support a simulation (Barsalou, Kyle Simmons, Barbey, & Wilson, 2003) of the likely surrounding world that the view only partially reveals. In real-world perception as visual sampling continues, the representation increasingly reflects the specific details of the surrounding scene. What is suggested here is that the first fixation on a scene is sufficient to initiate a simulation that subsequent views can confirm or correct and embellish. According to the multisource model, just as the visual field itself is inhomogeneous, scene simulation too is inhomogeneous, shading from the highly detailed visual information in the current view to the periphery of vision, to amodal perception just beyond the boundaries, and to increasingly general and schematic expectations. A key impetus for my colleagues and me in considering this alternative conceptualization of scene perception has been *boundary extension*, a constructive error in memory for views of the world that was first reported in Intraub and Richardson (1989).

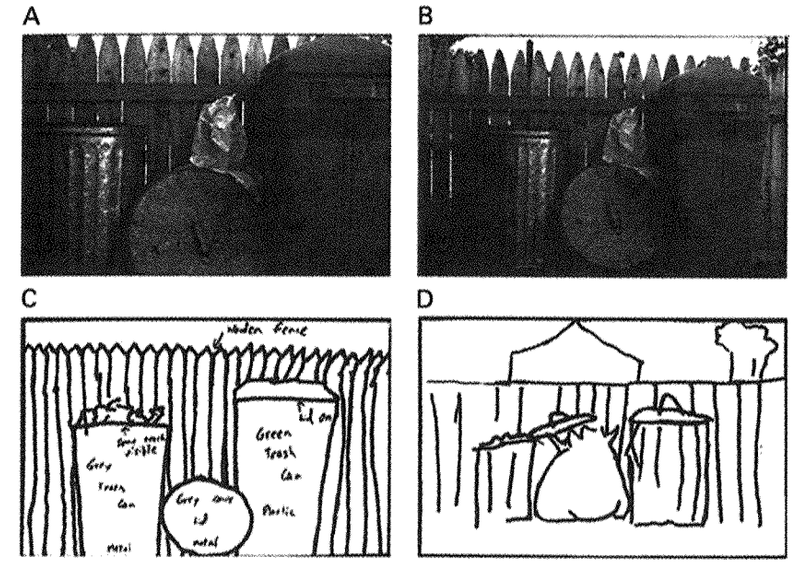

Figure 1.1
A pair of photographs (close-up and more wide-angle view) and a drawing of each one made by different participants. Note the content at the edges of the photographs and the edges of the drawings; participants remembered seeing more of the world than was shown (boundary extension). Based on figure 1.1 in Intraub and Richardson (1989), *Journal of Experimental Psychology: Learning, Memory & Cognition.*

Boundary Extension

Boundary extension (Intraub, 2010, 2012) is an error in which the observer remembers having seen beyond the physical boundaries of a view. Figure 1.1 shows an example of boundary extension in participants' drawings of photographs from memory (Intraub & Richardson, 1989). As shown in the figure, their drawings included more of the scene than was shown in the photograph. When the close-up view (panel A) was drawn from memory, not only did participants remember seeing complete garbage cans, but they remembered seeing a continuation of the fence beyond each one as well as more of the world beyond the upper and lower boundaries (as shown in the example in panel C). Although this overinclusive memory was an error with respect to the stimulus view, a comparison of this drawing with the wider-angle photograph of the same scene (panel B) shows that this error was also a good prediction of the world just beyond the boundaries of the original view. This effect was first discovered in the context of long-term memory (retention intervals of minutes to days: Intraub, Bender, & Mangels, 1992; Intraub & Richardson, 1989), but subsequent research has shown that boundary extension can occur across a masked retention interval as brief as 42

ms (commensurate with a saccade) as well as across an actual saccade when the stimulus and test picture fall on different sides of the screen (Dickinson & Intraub, 2008; Intraub & Dickinson, 2008). The rapid presence of extrastimulus scene layout is challenging to explain in terms of visual memory alone.

Boundary extension may not occur for all types visual stimuli (e.g., an object on a blank background) or at least not to the same degree (Gottesman & Intraub, 2002; Intraub, Gottesman, & Bills, 1998). It appears to be strongly associated with perceiving or thinking about scenes. For example, Intraub et al. (1998) presented line drawings of single objects on blank backgrounds. Participants were instructed to remember the objects and their sizes, but one group was induced to imagine a specific real-world location in the blank background. Although the visual information was the same, boundary extension occurred only in the scene imagination condition. An imagery control condition revealed that it was not imagery per se that had caused boundary extension in the "imagine-background" condition because when participants were instructed to imagine the colors of the objects during presentation, again boundary extension did not occur. When, in another condition, a simple background was added to each line drawing (consistent with the imagery inducement descriptions), again boundary extension occurred. In fact, performance was virtually identical when the background was imagined as when it was perceived. This, and related observations (Gottesman & Intraub, 2002) suggested that boundary extension is associated with the recruitment of processes associated with perceiving or thinking about spatial layout and meaningful locations (i.e., scenes).

Different aspects of spatial layout have been shown to affect boundary extension. The scope of the view is one such factor. Boundary extension is greatest in the case of tight close-ups and decreases, ultimately disappearing, as the view widens (Intraub et al., 1992). Close-ups yield the largest error even when there is a clear marker for boundary placement such as when a boundary slightly crops the main object in a picture. In an eye-tracking study, Gagnier, Dickinson, and Intraub (2013) presented participants with close-up photographs of single-object scenes either with or without a cropped object. Participants fixated the picture boundaries and, in the case of the cropped objects, fixated the place where the boundary cut across the object, and they refixated those areas, minutes later, at test. Yet, when participants adjusted the boundaries using the mouse to reveal more or less of the background at each picture boundary, they moved the boundaries outward to reveal unseen surrounding space in spite of their well-placed eye fixations.

To determine if knowing what would be tested in advance would influence oculomotor activity and eliminate boundary extension, another group of participants was forewarned about the nature of the test at the start of the experiment. Fixations to the boundary region and to the cropped region increased for this group, indicating that participants were attending to these critical areas. However, at test, although the

size of the boundary error was reduced, boundary extension again occurred. In spite of numerous fixations to the location where the boundary had cropped the object, participants moved the cropping boundary outward such that they not only completed the object, but showed additional background beyond. Thus, knowing what would be tested in advance, increasing eye fixations to the most informative areas during encoding (including a clear marker of boundary placement) and then fixating those regions again at test were not sufficient to overcome boundary extension. Memory was not constrained to the high-acuity visual information available during multiple fixations but also included additional nonvisual information consistent with representation of the view within a larger spatial context.

Spatial factors such as the distance of an object to a boundary (Bertamini, Jones, Spooner, & Hecht, 2005; Intraub et al., 1992) impact boundary extension. But in the context of the current discussion, it is important to note that this effect is not tied solely to distance *within* the picture space (e.g., "the object is 1 cm from the picture's left boundary") but reflects how much of the real world scene is depicted within that distance in the picture (e.g., how much of the background can be seen in that 1-cm space). Gagnier, Intraub, Oliva, and Wolfe (2011) kept the distance between the main object and the edges of the picture constant, but across conditions they changed the camera's viewpoint (0° degrees, straight ahead; or 45° angle, at an angle) so that the picture would include more or less of the scene's background within the space between the object and the picture's boundaries. Results showed that in this case, although the distance between the object and the picture's boundaries was the same, boundary extension was affected by how much of the world could be seen within that space. They found that at a 45° angle, with more of the background space visible than at the 0° angle (straight ahead), boundary extension was attenuated (similar to the reduction in boundary extension between more wide-angle and more close-up views).

The observations described to this point demonstrate that scene representation is not simply a visual representation of the photographs presented in these experiments but instead draws upon other sources of information about the likely surrounding scene that the photograph only partially revealed. Further support for this point is the observation that boundary extension occurs even in the absence of vision, when blindfolded participants use haptic exploration to perceive and remember objects that are meaningfully arranged on a natural surface (e.g., a place setting on a table: Intraub, 2004; also see Intraub, 2010). In Intraub (2004) each stimulus region was bounded by a wooden frame that limited the haptic participants' exploration. The frame was removed, and minutes later participants reconstructed the boundaries. Results revealed that these participants remembered having felt the "unfelt" world just beyond the edges of the original stimulus. To determine if this "haptic" boundary extension may have been mediated by visual imagery, in the same study a "haptic expert," a woman who had been deaf and blind since early life, explored the same

scenes. She too increased the area of the regions to include more of the background than she had actually touched.

In closing this section I should point out that although most research on boundary extension has been conducted with young adults, research thus far indicates that it occurs throughout the life span. Boundary extension has been reported in children's memory (4–10 years of age; Candel, Merckelbach, Houben, & Vandyck, 2004; Kreindel & Intraub, 2012; Seamon, Schlegel, Hiester, Landau, & Blumenthal, 2002), in older adults (Seamon et al., 2002), and in infants as young as 3 months of age (Quinn & Intraub, 2007). In sum, participants remembered having seen, felt, and imagined more of a view than was physically presented, even when they fixated the region near the boundary (visual exploration) or touched it (haptic exploration). In vision, they failed to recognize the identical view across a retention interval lasting less than 1/20th of a second. Intraub and Dickinson (2008) proposed a framework they referred to as the *multisource model* of scene perception that offers an explanation for these observations.

A Multisource Model of Scene Perception

As discussed earlier, Intraub and Dickinson (2008; Intraub, 2010, 2012) proposed that visual scene representation draws on multiple sources of top-down information in addition to the visual input. A depiction of the model is presented in figure 1.2. The construction of a multisource scene representation is depicted in the top panel of figure 1.2. The visual input is organized within the observer's spatial framework along with amodal perception beyond the boundaries and expectations and constraints based on rapidly available scene knowledge. In the case of a photograph, the observer takes the viewpoint of the camera (e.g., "in front of me" in typical photographs; "below me" in the case of a bird's-eye view). In fact, in photography, viewpoints such as low-angle, eye-level, and high-angle have been shown to influence the observer's interpretation of characters and events (Kraft, 1987). This organized multisource representation can be thought of as a mental *simulation* of the world that the visual information only partially reveals (Barsalou et al., 2003).

The top panel of figure 1.2 shows that this occurs while the sensory information is available. The presentation of a view elicits top-down sources of information about the larger scene (the likely surrounding world associated with the current view). All sources of information are available. While the sensory information is present, the dividing line between the sensory input and the top-down continuation of the scene is very clear. The observer can readily *see* the boundary of the view, even while thinking about the surrounding context. However, once the sensory input is gone, as in the lower panel of figure 1.2, what is now available is a remembered scene representation in which different parts were originally derived from different sources. Following the

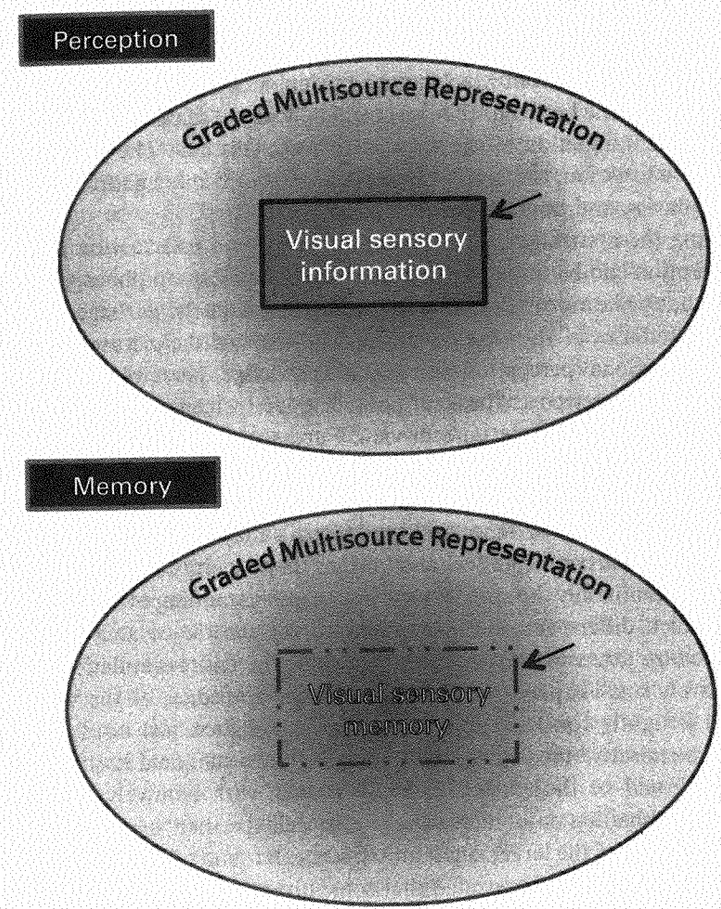

Figure 1.2
An illustration of how the multisource model accounts for boundary extension. In stage 1 (top panel), the sensory input, along with multiple top-down sources of information, creates a simulation of the likely surrounding world; the dividing line (designated by the arrow) between visible information in a photograph and the top-down continuation of the scene is easy to discern. In stage 2 (bottom panel), after the sensory input is gone, there are no tags to specify source; the dividing line (designated by the arrow) between visual memory for the once-visible information and memory for top-down generated information just outside the view is no longer clear. Information from just beyond the original boundaries was so well constrained by the visual input that in memory it is misattributed to a visual source, resulting in boundary extension.

key insight raised by Marcia Johnson and her colleagues (Johnson, Hashtroudi, & Lindsay, 1993; Johnson & Raye, 1981), the representation does not contain "tags" to indicate which parts were derived from which source (vision, amodal perception, and so forth). Now the dividing line between what was originally visual sensory input and originally the top-down continuation of the scene is no longer distinct. The observer may falsely misattribute to vision, the highly constrained expected information from just beyond the boundaries, and boundary extension is the result.

The idea that during the first stage several sources of top-down information contribute to scene perception can be thought of as follows. In the close-up photograph in figure 1.1A, note that while viewing the picture one can see that the garbage cans are cropped by the boundaries of the view on the left and right of the picture yet at the same time (through amodal perception and object knowledge) perceive the pails as whole. The pails are not perceived as "broken" pails. They are perceived to be whole pails that are just not fully visible in the current view. Furthermore, identification of the view as an "outdoor scene" carries with it clear implications that there must be a sky above and a ground plane below (even if these are not visible), and if the observer lives in the United States, the type of fence and type of pails may further specify an "outside area" that is "a suburban neighborhood" rather than "a city scene."

The simulation includes not only the studied view but an understanding of the likely surrounding scene, which differs across observers based on experience and world knowledge. The boundary extension error does not include the entire simulation; it involves only a relatively small expanse of space just beyond the edges of the view. Only the region that is tightly constrained by the visual information just inside the boundary is likely to be misattributed to vision. Again, like the visual field itself, this scene representation should be thought of as being graded, with relatively highly detailed visual memory shading into increasingly less well-specified expectations about the layout and content of the larger surrounding scene. Bar and colleagues (Bar, Aminoff, & Schacter, 2008) have demonstrated that some objects elicit a very specific surrounding context whereas others do not (and instead can be tied to multiple possible locations). Thus, different views of the world may yield very specific or very vague expectations about the surrounding world (including expectations about what one is most likely to see if one *could* turn one's head to see it).

In this account, boundary extension occurs very rapidly because the simulation is not generated *after* the stimulus is gone. What will become the boundary-extended region in memory is part of the scene simulation that becomes active within the first fixation. Once the stimulus is gone, even for a fraction of a second (Dickinson & Intraub, 2008; Intraub & Dickinson, 2008), the observer may misattribute to vision a small swath of surrounding space just beyond the original boundaries. The mistaken region is misattributed only because it so closely resembles the remembered visual information just inside the view. In the example in figure 1.1, completion of the

garbage pails and continuation of the fence are so highly constrained by the visual information that they are readily misattributed to having been seen. This, of course, is a theoretical hypothesis, but there are some observations in the literature that are consistent with this possibility..

In the source-monitoring framework (Johnson et al., 1993), the decision about the source of a memory is affected by the qualities of the remembered information. In the case of boundary extension, factors that increase the similarity between memory for the visually presented information just inside the boundary and the imagined continuation of that information just outside the boundary should therefore affect how much imagined space will be misattributed to vision. Consistent with this idea, Intraub, Daniels, Horowitz, and Wolfe (2008) found that when participants viewed 750-ms photographs under conditions of divided attention, which would be expected to compromise the quality of the visual input, they experienced *greater* boundary extension than when attention was not divided. Gagnier and Intraub (2012) found that memory for line drawings of complex scenes led to greater boundary extension than memory of color photographs of the same scenes. They suggested that the mental representation was more similar across the boundary for the simple lines in the line drawing than for the more complex visual information in the photograph. It is difficult to argue that divided attention would cause greater computation of surrounding space or that line drawings would evoke a greater sense of a specific surrounding world than would naturalistic photographs. Instead, the authors proposed that in all cases, the view rapidly activated a representation of the likely surrounding spatial layout, but that the attribution of source (seen vs. imagined) differed. Divided attention and simple line stimuli in the two examples just described may have helped to increase that similarity and thus led to more of the imagined surrounding space being misattributed to vision (i.e., a greater boundary extension error).

The boundary extension error itself has been described as an adaptive error in that predicting upcoming layout might facilitate view integration as we sample the world (Intraub, 1997). Evidence for its presence across a saccade (Dickinson & Intraub, 2008; Intraub & Dickinson, 2008) suggests that it is at least available to support integration of successive views. More direct evidence of its potential impact has been provided by Gottesman (2011) in a priming task. Using a modified version of Sanocki and Epstein's (1997) layout-priming paradigm, she demonstrated that, when boundary extension occurs, the falsely remembered region beyond the boundary can prime visual perception of that region when it is visually presented later. Perhaps similar priming occurs in the haptic modality, but as yet this has not been tested.

In evaluating the first stage of the model (generation of the multisource scene representation), is there any evidence to support the idea that a view elicits a

representation of surrounding space? In the next sections I describe behavioral evidence (descriptions of remembered views) and neuropsychological evidence for this proposition, and then I describe neuroimaging studies that suggest the neural architecture that may underlie both scene representation and boundary extension.

Scene Simulation: Evidence from Scene Descriptions

In day-to-day scene perception we are embedded within the scenes we perceive. What is suggested here is that the mental representation of a scene may reflect this physical reality (see Shepard, 1984, for a discussion of internalized constraints in perception). In Intraub (2010) I described an anecdotal illustration of scene simulation in which observers' interpretation of a scene had clearly drawn on expectations that went well beyond the visual information in the picture. The picture that elicited this simulation was the mundane photograph of garbage cans from Intraub and Richardson (1989) shown in figure 1.1. I had always interpreted this photograph as depicting garbage awaiting pickup on the street. I thought that my co-author, Mike Richardson, had set the tripod in a suburban street with another neighbor's house behind him. In fact, for years, I admonished students not to stand in the street when taking pictures, for safety's sake. When more recently I asked a colleague if from memory he had a sense of the camera's location and what was behind the photographer, he quickly reported that the garbage was in a backyard and that the photographer was standing with his back to the owner's house. To my surprise, another colleague, when asked, immediately said that the photographer was in an alley with the other side of the alley behind him and added, "Where else would he be?" After some confusion (for me, "alley" brought to mind a dark New York City–style alley between large buildings), she explained that in the southwestern United States where she had spent most of her life, suburban streets are aligned such that backyards on adjacent streets abut a street that serves as an alleyway behind the houses for garbage pickup.

Recently I contacted six researchers[1] in different regions of the United States and one from the United Kingdom who I thought would have long-term memory for this photograph, and asked them the same questions. Their responses are shown in table 1.1. Respondents 1–5 are from the United States (respondent 2 is from the Southwest). Of interest is the very different response offered by respondent 6, who remembered the picture but clearly had no sense of a locale. He offered only that the camera was positioned in front of the garbage cans. Further inquiry revealed that where he lives, these types of receptacles (particularly the metal can) are atypical and that he associates them mostly with old U.S. cartoons. Thus, the context for him was *weak* without a strong sense of locale. This may be an example of the observation by Bar et al. (2008) that different objects can elicit either strong or weak contexts, in this case specifically tied to this respondent's experience with the objects.

Table 1.1
Answers to where the camera was and what was behind it based on six researchers' long-term memory for the photograph of garbage cans by a fence (see figure 1.1)

Respondent	Response to location of the camera and what was behind the photographer.
1	I've always thought of it as a scene at the side fence of a house accessible to the front (so that the garbage collectors can get to the cans) but not in plain view when you're out in the backyard.
2	Photographer was standing in an alley (extending to the left and right). Behind him/her was another fence, and beyond that, another house.
3	[The camera is] in front of the fence, as if the photographer was standing at the back of a house looking into the backyard.
4	The cans were against a wooden picket fence, so I assumed that the photographer was on the far side of a driveway or possibly a small parking area.
5	The photographer was probably standing on the street.
6	[After providing description of objects in the picture] ... I guess the photographer must have been straight in front of the fence. No idea what was behind him.

Note: Respondents 1–5 are based in the United States; respondent 6 is based in the United Kingdom (see text).

These reports are anecdotal, but they suggest that even a view as mundane as in the photos in figure 1.1 can evoke a representation of a coherent surrounding world—especially when familiar objects are presented. Unbidden, specific surroundings came to mind and appear to have been part of the interpretation of the view (e.g., garbage awaiting pickup on the street or garbage in a backyard or garbage in an alleyway). An interesting aspect of the reports is the commitment to a particular locale in the participant's mind that the view evoked. Those who were queried claimed to have always "thought of the picture this way." There is no reason to think that any of these different mental scenarios would impact boundary extension (how much extrastimulus information they later attributed to vision) as, in all scenarios, immediately beyond the edges of the view there is likely to be more of the fence and a continuation of the background above and below the given view. However, the commitment to a locale and an imagined surrounding world raise the question of whether observers who have a deficit in imagining a surrounding world might also be prone to little or no boundary extension.

A Neuropsychological Approach to Scene Simulation and Boundary Extension

Hassabis and colleagues (Hassabis, Kumaran, Vann, & Maguire, 2007) reported an interesting, previously unknown deficit associated with bilateral hippocampal lesions. In addition to the expected memory deficit of anterograde amnesia, all but one of the patients tested (a patient with some spared hippocampal tissue) also exhibited a markedly impaired ability to imagine a coherent surrounding scene that they were asked to create in response to a verbal cue (e.g., "Imagine you are lying on a white sandy beach in a beautiful tropical bay"). Ten such scenarios (referring to common locales,

such as visiting an art museum or market, or self-referential future scenarios, such as imagining a possible event over the next weekend) were presented, and both the patients and the matched controls were encouraged to "give free reign to their imaginations and create something new." They were encouraged to think of themselves as physically present in the surroundings and were asked to describe as many sensory details and feelings as they could.

Patients' descriptions seemed to lack spatial coherence. Their imagined worlds were fragmented and lacking in detail. It is important to note that the patients could report appropriate objects that matched the semantic context of the specified scenario, but spatial references that were apparent in the control participants' descriptions (e.g., "behind me is a row of palm trees....") were lacking in the patients' descriptions. The content of their descriptions and their subjective reports about the problems they encountered in trying to imagine a coherent world differed markedly from that of the matched control participants. Hassabis, Kumaran, Vann, et al. (2007; see also Hassabis, Kumaran, & Maguire, 2007) suggested that underlying the ability either to reconstruct a scenario from one's past or to imagine a new one (in one's future, or simply a new event based on one's general knowledge) relies on the ability to maintain a coherent, multimodal spatial representation of the event.

If we consider the multisource model depicted in figure 1.2, how might a lack of a spatially coherent scene simulation in the first stage impact boundary memory later? Mullally, Intraub, and Maguire (2012) sought to determine if patients with bilateral hippocampal lesions would be more resistant to boundary extension than their matched control participants. In terms of the multisource model, if the surrounding context lacks spatial coherence and detail, then very little if any of that imagined representation will be misattributed to vision after the stimulus is gone. Paradoxically, this hypothesis predicts that patients who suffer from severe memory deficits would actually have a more veridical memory for views of the world than would their matched controls. To test this hypothesis, Mullally, Intraub, and Maguire (2012) chose three different protocols for assessing boundary extension that would fall within the patients' memory span. The first was a brief presentation paradigm (Intraub & Dickinson, 2008), the second was an immediate drawing task (Kreindel & Intraub, 2012; Seamon et al., 2002), and the third was a haptic border reconstruction task (Intraub, 2004).

To ascertain if this group of patients showed the same scene construction deficits as in Hassabis, Kumaran, Vann, et al. (2007), a similar set of scene construction tasks was also administered. As in that study, the assessment revealed a deficient ability to construct a spatially coherent imagined world. Patients offered such comments as the imagined space being "squashed," and they provided fragmented descriptions. A new scene probe task was developed in which the patients looked at photographs and were asked to describe what they thought would be likely to exist just beyond the

boundaries of the view. They did not differ from the controls in naming semantically appropriate objects, sensory description, or thoughts, emotions, or actions. However, they produced significantly fewer spatial references in describing the content. Given these problems, how did they fare on the boundary extension tasks?

In the brief presentation task, on each trial they were presented with a photograph for 250 ms. The view was interrupted by a 250-ms mask and then reappeared and stayed on the screen. The participant then rated the test view as being the same, closer up (bigger object, less surrounding space) or farther away (smaller object, more surrounding space) than before on a five-point scale. In all cases, the picture following the 250-ms masked retention interval was identical to the stimulus view. Boundary extension occurred in both groups, but was greater in the control group. Figure 1.3 shows the number of times participants in each group classified the same view as "more close up," "the same," and "farther away." As the figure shows, control participants were more likely to erroneously rate the identical test view as looking "too close up" (indicating that they remembered the view before the mask as having shown more of the scene). Patients were more accurate in recognizing that the views were actually the same. Classifying the test views as "farther away" (smaller object, more surrounding space) was relatively rare and did not differ between groups. Thus, patients did not appear to be randomly selecting responses. They appeared to be more accurate in recognizing identical views after a 250-ms masked retention interval. This better accuracy was mirrored in the pattern of confidence ratings. Control participants were

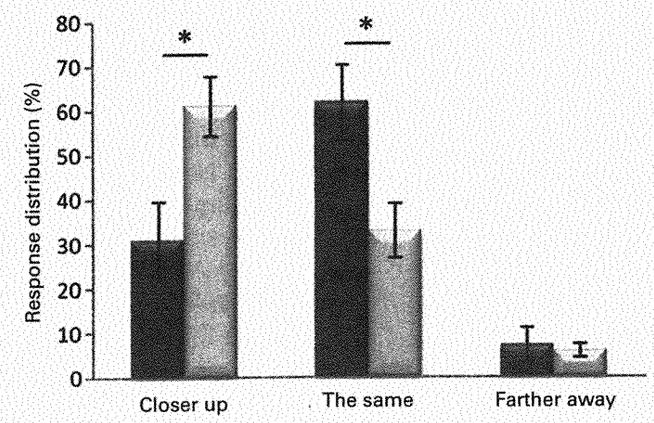

Figure 1.3
The proportion of trials classified as either "closer up," "the same" (correct answer), or "farther away" was calculated and represented as a percentage response distribution score for the patients (bilateral hippocampal lesions) and their matched control participants. Reproduction of a panel presented in figure 1.2 of Mullally, Intraub, and Maguire (2012), *Current Biology*.

more confident in their erroneous boundary extension responses than in their correct "same" responses, whereas patients were more confident of their same responses than their erroneous boundary extension responses.

In the drawing task participants viewed a photograph for 15 seconds and then immediately drew it from memory. Both the patients and their matched control participants drew boundary-extended pictures. They reduced the size of the main object and included more surrounding background in their drawings than was shown in the photograph they had just studied. However, again, patients exhibited less boundary extension than did their matched control participants. Their drawings more accurately captured the view. The three photographs that served as stimuli and the drawings made by one patient and her matched control participants are shown in the upper panel of figure 1.4. In the lower panel the graph shows the reduction in size of the main object in the drawing (as compared with the object in the photograph) for the patient and her control participants. The patients' objects and the amount of visible background drawn were more similar to the view in the original photograph than were those of their matched control participants. A group of independent judges rated all the drawings in the study and found no difference in detail or quality between the pictures drawn by patients and those drawn by control participants. A separate group of independent judges could not discriminate which pictures were drawn by patients versus control participants. The patients simply appeared to be more resistant to the boundary extension error.

In the third task haptic exploration of objects in settings similar to those in Intraub (2004) was undertaken by both groups. Participants were blindfolded and felt the objects and backgrounds of small scenarios bounded by a wooden frame. After they had explored each scenario, the frame was removed, and the participants immediately reexplored the region, indicating where each boundary had been located. In this case the control participants showed significant boundary extension, setting the boundaries outward; on average they increased the total area by about 12%. No reliable change in position was observed for the patients, the *N* was small, but the direction of the mean area remembered did not suggest boundary extension (it was reduced by about 5%). In sum, across all three tasks, patients who showed poor spatial coherence in scene construction and imagination tasks were also more resistant to boundary extension than were their matched controls.

Neuroimaging and Boundary Extension

The first report of the parahippocampal place area (PPA) (Epstein & Kanwisher, 1998) was published the same year as Intraub et al. (1998). What was striking to us was that Epstein and Kanwisher had found that pictures of locations (e.g., the corner of a room) caused heightened activation in PPA, whereas objects without scene

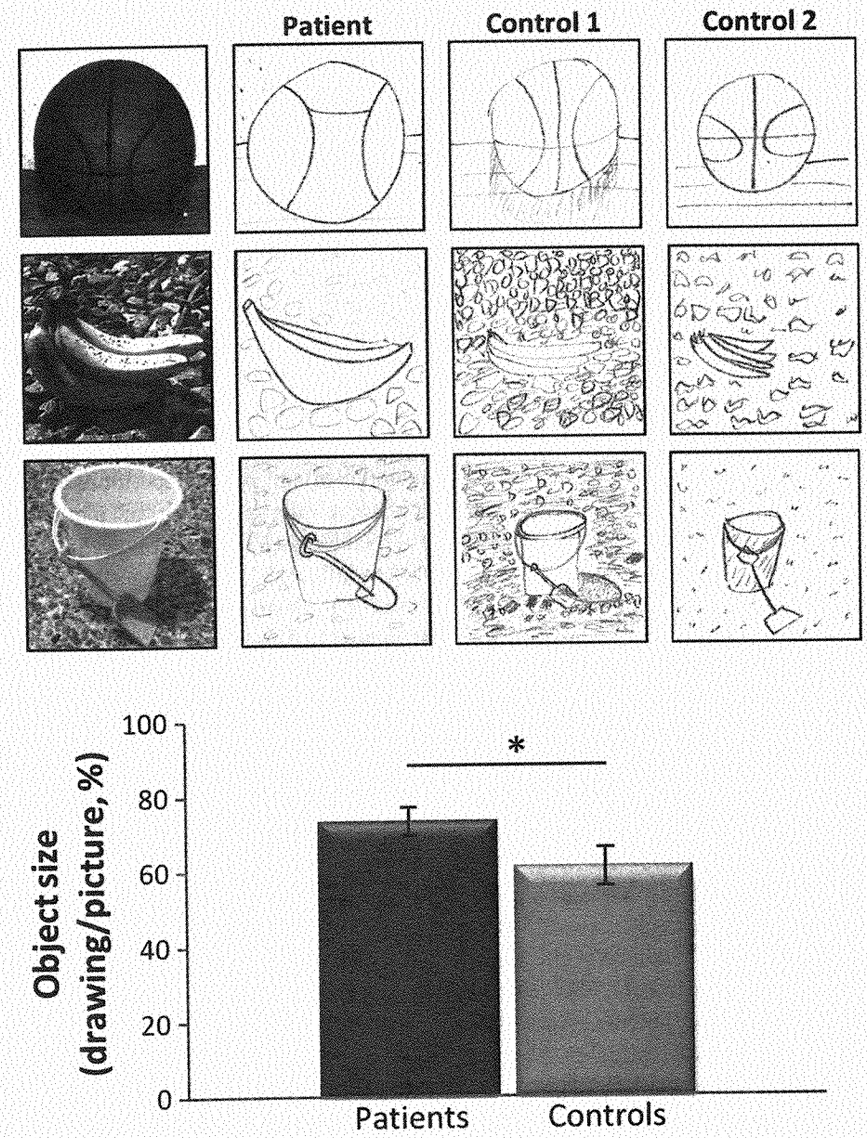

Figure 1.4
The left panel displays the three scene stimuli. Sample drawings by a patient and her two matched control participants are displayed in the middle and left panels. The graph shows that the control subjects reduced the size of the main object, incorporating more background in their drawings than did the patients. Data are presented as means +1 SEM; $*p < 0.05$. Reproduction of figure 1.3 of Mullally, Intraub, and Maguire (2012), *Current Biology*.

context did not, and we had found that outline drawings of objects in a scene location elicited boundary extension, whereas the same outline drawing of objects on a blank background did not. O'Craven and Kanwisher (2000) reported that simply imagining a location was sufficient to increase PPA activation, and Intraub et al. (1998) found that drawings of outline objects on blank backgrounds would result in boundary extension if the observer imagined a real-world location filling the blank background. Because both boundary extension and activation in PPA seemed to be tied in some way to layout and scene representation, we wondered if boundary extension might be associated in some way with the PPA.

This possibility was subsequently explored in an fMRI experiment (Park, Intraub, Yi, Widders, & Chun, 2007) in which a behavioral pattern of errors that is diagnostic of boundary extension was exploited. Pairs of closer-up and wider-angle views of single object scenes were used in the experiment. Each picture was presented for a few seconds at a time in a series. At various lags, the scene repeated, either as the identical view or as a mismatched view (i.e., the other member of the pair). Repetitions that were mismatched were of particular interest because the same pair of pictures was presented in both cases, just in a different order (the closer-up view followed later by the wider-angled view or the wider-angled view followed later by the closer-up view). In behavioral studies (beginning with Intraub & Richardson, 1989), when participants rated a mismatched test picture on the five-point boundary scale described earlier, a marked asymmetry was observed. When the order was *close-then-wide*, participants rated the mismatched view as being more similar to the original view than when the order was *wide-then-close*. Presumably this is because boundary extension in memory for the first picture caused it to more closely match perception of the second picture in the *close-wide* case, whereas it exaggerated the difference between pictures in the *wide-close* case.

When fMRI data were recorded, the participants simply watched the stimuli, making no behavioral responses at all. They were instructed to try to remember the pictures in as much detail as possible (focusing on both the object and the background) and were informed that the same scenes would sometimes repeat. Analysis of adaption responses in both PPA and the retrosplenial complex (RSC) revealed the diagnostic asymmetry. The neural response to the second picture was attenuated in the *close-wide* case (suggesting that to these regions the stimuli were very similar), whereas the neural response to the second picture showed no attenuation in the *wide-close* case (suggesting that to these regions the stimuli were quite different). However, in lateral occipital cortex (associated with object recognition, but not the size of the object) attenuation occurred in both cases (to this region the stimuli were the same regardless of the view). The pattern of neural attenuation in PPA and RSC suggested that both areas were sensitive to the boundary-extended representations of the

pictures rather than to the physical views that were presented. Following the fMRI study, the same participants took part in a behavioral boundary extension experiment, and their explicit boundary ratings also revealed the typical asymmetry.

Epstein (2008, 2011; Epstein & Higgins, 2007) has proposed that PPA and RSC are part of the neural architecture that underlies navigation and integration of views within larger contexts (although Bar et al., 2008, have suggested that these regions may also be involved in conceptual processing of contextual associations; please see chapter 6 by Epstein and chapter 7 by Aminoff in this volume). Most research on scene representation that explores these ROIs has made use of visual stimuli, but research contrasting neural responses in PPA and RSC to haptic exploration of Lego scenes in contrast to Lego objects (no scene context) by blindfolded participants, as well as by congenitally blind participants, supports the idea that these areas may be responding to the spatial structure that underlies both visual and haptic exploration (Wolbers, Klatzky, Loomis, Wutte, & Giudice, 2011). Epstein (2011), in a discussion of the implications of the Wolbers et al. finding, drew the comparison to boundary extension, which, as described earlier, occurs whether exploration is visual or haptic (when the latter condition included either sighted participants who were blindfolded or a woman who had been deaf and blind since early life; Intraub, 2004). These results provide support for the value of moving from a modality-centric view of scene perception to a spatio-centric conceptualization.

The Park et al. (2007) study addressed attenuation of responses in PPA and RSC, thereby focusing on memory (the second stage of the multisource model). It cannot provide insight into brain activity associated with the first stage of the model—generation of a scene simulation. The Mullally et al. (2012) study of boundary extension in patients with bilateral hippocampal lesions suggested a possible role of hippocampus in developing a coherent spatial representation of a scene. Chadwick, Mullally, and Maguire (2013) conducted an fMRI experiment to test this possibility. They used a modified version of Intraub and Dickinson's (2008) brief presentation task but on all trials presented a close-up tested by the identical close-up (timing was similar to that in the brief presentation Experiment in Mullally et al., 2012). In brief-presentation boundary extension experiments (unlike those using longer multisecond presentation), although boundary extension occurs overall, on many trials ratings suggest no boundary extension; for example, in Intraub and Dickinson (2008) boundary extension occurred for the majority of pictures (64%) but not for all pictures. Chadwick et al. (2013) cleverly exploited this and compared the neural response in hippocampus in trials in which boundary extension occurred and trials in which it did not (based on the participants' behavioral rating in each trial).

In line with Chadwick et al.'s hypothesis, greater hippocampal activation was associated with stimuli in trials in which the behavioral response to the test picture was

consistent with boundary extension than in those in which it was not. They reported that this neural response occurred before the onset of the test stimulus (the retention interval ranged from 1.2 to 3.2 seconds), so this response could not be attributed to the memory error at test.

Chadwick et al. also analyzed habituation responses in PPA (here referring to specific regions of PHC), RSC, and visual cortex. PHC and RSC were sensitive to the boundary extension error (as in the Park et al., 2007, study). Greater adaptation occurred in trials during which participants had correctly identified the views as being the same, and less adaptation occurred in trials during which they did not, and a similar adaptation effect was observed in visual cortex. What is important to remember is that the stimulus and test pictures in the experiment were always identical, so the differences observed were not mediated by any visual differences between stimulus and test. Chadwick et al. conducted DCM connectivity analyses that suggested that the hippocampus was driving the responses in the other regions of interest (no habituation effects were observed in the hippocampus). They concluded by suggesting that the neural responses observed are consistent with the two-stage model of boundary extension in which the first stage involves computation of a spatially coherent scene representation and the second stage involves the boundary error (PHC and RSC are sensitive to that *boundary-extended* representation).

The early hippocampal response (putatively tied to the first stage) in conjunction with the adaptation responses (in PHC and RSC) led Chadwick et al. to suggest that the hippocampus plays a fundamental role in supporting construction of a spatially coherent scene representation that is "channeled backwards through the processing hierarchy via PHC and as far as early visual cortex" to provide predictions about the likely surrounding world. This parallels the first stage of processing described earlier in figure 1.2. Subsequently, adaptation responses arise in response to a representation that now includes extended boundaries. The neuroimaging data in combination with the neuropsychological data suggest that the hippocampus might be involved in scene construction when the observer is presented with a view of the world. It is interesting that other research has suggested a role for the hippocampus not only in supporting episodic memory (and reconstruction of past events) but in supporting simulation of future events (e.g., future planning: Addis, Cheng, Roberts, & Schacter, 2011; Hassabis & Maguire, 2007). It is suggested here that scene perception itself (within the present moment as we sample the world around us) may involve many of the same simulation processes. The behavioral and neuroimaging research on boundary extension reviewed here suggests that the traditional modality-centric approach to scene perception does not capture the complexity of what it means to understand a scene and that a spatiocentric approach provides a viable alternative that would incorporate the visual input within a multisource cognitive representation of surrounding space.

Conclusion

Boundary extension provides an unusual means for exploring scene representation because people remember having seen beyond the boundaries of the physical view. They remember experiencing a region of space in the absence of any corresponding sensory input. Participants include this unseen region in their drawings, they move the boundaries outward to reveal this space in interactive border adjustment tasks, and they rate the scope of the view incorrectly, indicating that they remember seeing more of the scene than was actually presented (Intraub, 2010). Neural responses in PPA and RSC reflect this same overinclusive memory for recently presented views (Chadwick et al., 2013; Park et al., 2007) in tasks that elicit boundary extension. The multisource model (Intraub, 2010, 2012; Intraub & Dickinson, 2008) can account for these observations by replacing a modality-specific framework of scene representation (e.g., a visual representation) with a multisource representation organized around the observer's sense of surrounding space. Recent neuropsychological and neuroimaging evidence have suggested that the hippocampus may play a role in the mental construction of this surrounding space (Chadwick et al., 2013; Mullally et al., 2012). This provides a potential bridge between research on scene perception and research on mental constructions (and their associated neural structures) that are thought to be involved in remembering past scenarios and in generating representations of future scenarios (Addis et al., 2011; Addis, Wong, & Schacter, 2007; Hassabis & Maguire, 2007; Johnson & Sherman, 1990). In the case of scene perception, the brain's ongoing constructive activity is focused not only on long-term memory or on distant future projections, but on the present, as we perceive and interact with our immediate surrounding world.

Note

1. Many thanks to Marco Bertamini, Tim Hubbard, Geoff Loftus, Greta Munger, Dan Reisberg, and Dan Simons for their descriptions.

References

Addis, D. R., Cheng, T., Roberts, R. P., & Schacter, D. L. (2011). Hippocampal contributions to the episodic simulation of specific and general future events. *Hippocampus*, *21*(10), 1045–1052.

Addis, D. R., Wong, A. T., & Schacter, D. L. (2007). Remembering the past and imagining the future: Common and distinct neural substrates during event construction and elaboration. *Neuropsychologia*, *45*(7), 1363–1377.

Bar, M. (2004). Visual objects in context. *Nature Reviews Neuroscience*, *5*(8), 617–629.

Bar, M., Aminoff, E. M., & Schacter, D. L. (2008). Scenes unseen: The parahippocampal cortex intrinsically subserves contextual associations, not scenes or places per se. *Journal of Neuroscience*, *28*(34), 8539–8544.

Barsalou, L. W., Kyle Simmons, W., Barbey, A. K., & Wilson, C. D. (2003). Grounding conceptual knowledge in modality-specific systems. *Trends in Cognitive Sciences*, *7*(2), 84–91.

Bertamini, M., Jones, L. A., Spooner, A., & Hecht, H. (2005). Boundary extension: The role of magnification, object size, context, and binocular information. *Journal of Experimental Psychology. Human Perception and Performance*, *31*(6), 1288–1307.

Candel, I., Merckelbach, H., Houben, K., & Vandyck, I. (2004). How children remember neutral and emotional pictures: Boundary extension in children's scene memories. *American Journal of Psychology*, *117*, 249–257.

Chadwick, M. J., Mullally, S. L., & Maguire, E. A. (2013). The hippocampus extrapolates beyond the view in scenes: An fMRI study of boundary extension. *Cortex*, *49*(8), 2067–2079.

Dickinson, C. A., & Intraub, H. (2008). Transsaccadic representation of layout: What is the time course of boundary extension? *Journal of Experimental Psychology. Human Perception and Performance*, *34*(3), 543–555.

Epstein, R. A. (2008). Parahippocampal and retrosplenial contributions to human spatial navigation. *Trends in Cognitive Sciences*, *12*(10), 388–396.

Epstein, R. A. (2011). Cognitive neuroscience: Scene layout from vision and touch. *Current Biology*, *21*(11), R437–R438.

Epstein, R. A., & Higgins, J. S. (2007). Differential parahippocampal and retrosplenial involvement in three types of visual scene recognition. *Cerebral Cortex*, *17*(7), 1680–1693.

Epstein, R. A., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature*, *392*(6676), 598–601.

Fantoni, C., Hilger, J. D., Gerbino, W., & Kellman, P. J. (2008). Surface interpolation and 3D relatability. *Journal of Vision*, *8*(7), 1–19.

Gagnier, K. M., Dickinson, C. A., & Intraub, H. (2013). Fixating picture boundaries does not eliminate boundary extension: Implications for scene representation. *Quarterly Journal of Experimental Psychology*, *66*, 2161–2186.

Gagnier, K. M., & Intraub, H. (2012). When less is more: Line-drawings lead to greater boundary extension than color photographs. *Visual Cognition*, *20*(7), 815–824.

Gagnier, K. M., Intraub, H., Oliva, A., & Wolfe, J. M. (2011). Why does vantage point affect boundary extension? *Visual Cognition*, *19*, 234–257.

Gottesman, C. V. (2011). Mental layout extrapolations prime spatial processing of scenes. *Journal of Experimental Psychology. Human Perception and Performance*, *37*(2), 382–395.

Gottesman, C. V., & Intraub, H. (2002). Surface construal and the mental representation of scenes. *Journal of Experimental Psychology. Human Perception and Performance*, *28*(3), 589–599.

Greene, M. R., & Oliva, A. (2009). The briefest of glances: The time course of natural scene understanding. *Psychological Science*, *20*(4), 464–472.

Hassabis, D., Kumaran, D., & Maguire, E. A. (2007). Using imagination to understand the neural basis of episodic memory. *Journal of Neuroscience*, *27*(52), 14365–14374.

Hassabis, D., Kumaran, D., Vann, S. D., & Maguire, E. A. (2007). Patients with hippocampal amnesia cannot imagine new experiences. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(5), 1726–1731.

Hassabis, D., & Maguire, E. A. (2007). Deconstructing episodic memory with construction. *Trends in Cognitive Sciences*, *11*(7), 299–306.

Hochberg, J. (1986). Representation of motion and space in video and cinematic displays. In K. J. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance* (Vol. 1, pp. 22.21–22.64). New York: Wiley.

Intraub, H. (1997). The representation of visual scenes. *Trends in Cognitive Sciences*, *1*(6), 217–222.

Intraub, H. (2004). Anticipatory spatial representation of 3D regions explored by sighted observers and a deaf-and-blind observer. *Cognition*, *94*(1), 19–37.

Intraub, H. (2010). Rethinking scene perception: A multisource model. In B. Ross (Ed.), *Psychology of learning and motivation: Advances in research and theory* (Vol. 52, pp. 231–264). San Diego: Elsevier Academic Press.

Intraub, H. (2012). Rethinking visual scene perception. *Wiley Interdisciplinary Reviews: Cognitive Science*, *3*(1), 117–127.

Intraub, H., Bender, R. S., & Mangels, J. A. (1992). Looking at pictures but remembering scenes. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *18*(1), 180–191.

Intraub, H., Daniels, K. K., Horowitz, T. S., & Wolfe, J. M. (2008). Looking at scenes while searching for numbers: Dividing attention multiplies space. *Perception & Psychophysics*, *70*(7), 1337–1349.

Intraub, H., & Dickinson, C. A. (2008). False memory 1/20th of a second later: What the early onset of boundary extension reveals about perception. *Psychological Science*, *19*(10), 1007–1014.

Intraub, H., Gottesman, C. V., & Bills, A. J. (1998). Effects of perceiving and imagining scenes on memory for pictures. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *24*(1), 186–201.

Intraub, H., & Richardson, M. (1989). Wide-angle memories of close-up scenes. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *15*(2), 179–187.

Irwin, D. E. (1991). Information integration across saccadic eye movements. *Cognitive Psychology*, *23*(3), 420–456.

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin*, *114*(1), 3–28.

Johnson, M. K., & Raye, C. L. (1981). Reality monitoring. *Psychological Review*, *88*, 67–85.

Johnson, M. K., & Sherman, S. J. (1990). Constructing and reconstructing the past and the future in the present. In E. T. Higgins & R. M. Sorrentino (Eds.), *Handbook of motivation and social cognition: Foundations of social behavior* (pp. 482–526). New York: Guilford Press.

Kanizsa, G. (1979). *Organization in vision*. New York: Praeger.

Kraft, R. N. (1987). The influence of camera angle on comprehension and retention of pictorial events. *Memory & Cognition*, *15*(4), 291–307.

Kreindel, E., & Intraub, H. (2012). *Boundary extension in children vs. adults: What developmental differences may tell us about scene representation.* Paper presented at the Vision Sciences Society, Naples, FL.

Mullally, S. L., Intraub, H., & Maguire, E. A. (2012). Attenuated boundary extension produces a paradoxical memory advantage in amnesic patients. *Current Biology*, *22*(4), 261–268.

O'Craven, K. M., & Kanwisher, N. (2000). Mental imagery of faces and places activates corresponding stiimulus-specific brain regions. *Journal of Cognitive Neuroscience*, *12*(6), 1013–1023.

O'Regan, J. K. (1992). Solving the "real" mysteries of visual perception: The world as an outside memory. *Canadian Journal of Experimental Psychology*, *46*(3), 461–488.

Park, S., Intraub, H., Yi, D. J., Widders, D., & Chun, M. M. (2007). Beyond the edges of a view: Boundary extension in human scene-selective visual cortex. *Neuron*, *54*(2), 335–342.

Potter, M. C. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology. Human Learning and Memory*, *2*(5), 509–522.

Quinn, P. C., & Intraub, H. (2007). Perceiving "outside the box" occurs early in development: Evidence for boundary extension in three- to seven-month-old infants. *Child Development*, *78*(1), 324–334.

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology (Hove)*, *62*(8), 1457–1506.

Rayner, K., & Pollatsek, A. (1992). Eye movements and scene perception. *Canadian Journal of Psychology*, *46*(3), 342–376.

Sanocki, T., & Epstein, W. (1997). Priming spatial layout of scenes. *Psychological Science*, *8*(5), 374–378.

Seamon, J. G., Schlegel, S. E., Hiester, P. M., Landau, S. M., & Blumenthal, B. F. (2002). Misremembering pictured objects: People of all ages demonstrate the boundary extension illusion. *American Journal of Psychology*, *115*(2), 151–167.

Shepard, R. N. (1984). Ecological constraints on internal representation: Resonant kinematics of perceiving, imagining, thinking, and dreaming. *Psychological Review*, *91*(4), 417–447.

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, *381*(6582), 520–522.

Tversky, B. (2009). Spatial cognition: Embodied and situated. In P. Robbins & M. Aydede (Eds.), *The Cambridge handbook of situated cognition* (pp. 201–216). Cambridge: Cambridge University Press.

Volkmann, F. C. (1986). Human visual suppression. *Vision Research*, *26*(9), 1401–1416.

Wolbers, T., Klatzky, R. L., Loomis, J. M., Wutte, M. G., & Giudice, N. A. (2011). Modality-independent coding of spatial layout in the human brain. *Current Biology*, *21*(11), 984–989.

Yin, C., Kellman, P. J., & Shipley, T. F. (2000). Surface integration influences depth discrimination. *Vision Research*, *40*(15), 1969–1978.