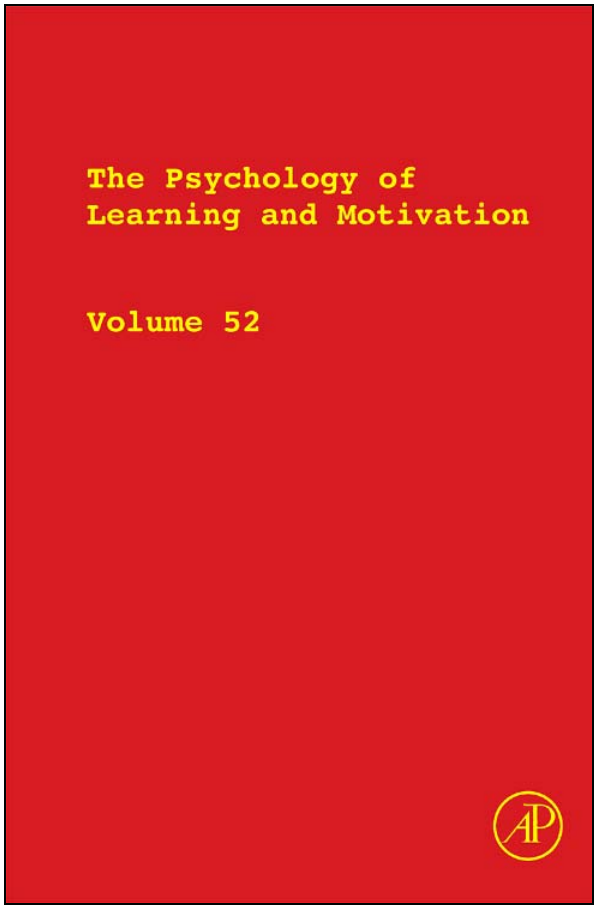

From: Helene Intraub, Rethinking Scene Perception: A Multisource Model. In Brian H. Ross, editor: The Psychology of Learning and Motivation, Vol. 52, Burlington: Academic Press, 2010, pp. 231-264.
ISBN: 978-0-12-380908-7

CHAPTER SIX

Rethinking Scene Perception: A Multisource Model

Helene Intraub

Contents

Abstract

Traditional approaches to scene perception begin with the visual input and track its progress through a series of very short-term memory buffers. While providing explanations for errors of omission (e.g., change blindness), such models are not as well suited for explaining rapid errors of commission, such as *boundary extension* [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*, 1007–1014]. I will present a multisource

Psychology of Learning and Motivation, Volume 52
ISSN 0079-7421, DOI: 10.1016/S0079-7421(10)52006-1

model of scene perception that begins instead with an egocentric reference frame. Even when the primary input is visual, the content that fills out this framework is derived from multiple sources (e.g., visual sensory, amodal, conceptual, and contextual). The multisource framework provides a novel explanation of boundary extension as a *source monitoring error* [Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin, 114*, 3–28] that arises when one attempts to discern which portion of the entire scene representation matches the visual sensory source. Behavioral and neuroimaging research with pictures and research on visual and haptic exploration of peripersonal space will be discussed. In the multisource view, scene perception is an act of spatial cognition that subserves many modalities—one of which is vision.

1. Introduction

We sit in rooms, walk through forests, work at desks, and cook in kitchens. How well can we remember specific views of the world we interact with every day? Research on visual scene perception and memory provides us not only with examples of stunning recognition memory performance for massive numbers of photographs (Standing, 1973; Standing, Conezio, & Haber, 1970), but also with stunning failures to recognize even sizeable changes in a photograph across a brief transient (Rensink, O'Regan, & Clark, 1997) or while a change slowly unfolds right in front of the viewer's eyes (Simons, Franconeri, & Reimer, 2000). Contrasts such as these have fueled debates about the extent to which we can retain the details of what we see (e.g., Henderson & Hollingworth, 2003; Rensink, 2000). *Boundary extension* (memory beyond the edges of a view; Intraub & Richardson, 1989; see Michod & Intraub, 2009 for an overview) raises a different set of challenges for understanding scene representation because it reflects neither the retention of detail nor the loss of detail; instead it is an error of *commission*. People remember seeing what *was not* visible, under conditions that are not expected to induce false memory.

Over the past 20 years, boundary extension has been reported under conditions that would normally be expected to support excellent memory, for example, low memory load (as few as 1–3 pictures; Bertamini, Jones, Spooner, & Hecht, 2005; Dickinson & Intraub, 2008; Intraub & Dickinson, 2008; Intraub, Gottesman, Willey, & Zuk, 1996; Intraub, Hoffman, Wetherhold, & Stoehs, 2006), distinctive pictures, and instructions that, following Intraub and Richardson (1989), are worded to draw as much attention to the background and layout of the picture as to the main object(s). Boundary extension has been reported for observers ranging in age from 6 to 84 years old (Seamon, Schlegel, Hiester, Landau, & Blumenthal, 2002, including children with Asperger's syndrome,

Chapman, Ropar, Mitchell, & Ackroyd, 2005). There is also evidence to suggest that infants as young as 3–4 months of age are subject to the same anticipatory spatial error (Quinn & Intraub, 2007).

How soon after the picture is gone does boundary extension occur? Generally speaking, errors of commission are thought to require heavy cognitive loads (e.g., long retention intervals, large stimulus sets, or stimuli that are confusable because they bear semantic or physical similarities) or inattention (cf. Koriat, Goldsmith, & Pansky, 2000). With this in mind, our first formal tests of boundary extension (Intraub & Richardson, 1989) were administered after relatively long retention intervals. The recall/drawing task was administered after retention intervals of 35 min or 2 days (e.g., see Figure 1, left column). Because of reports of excellent picture recognition

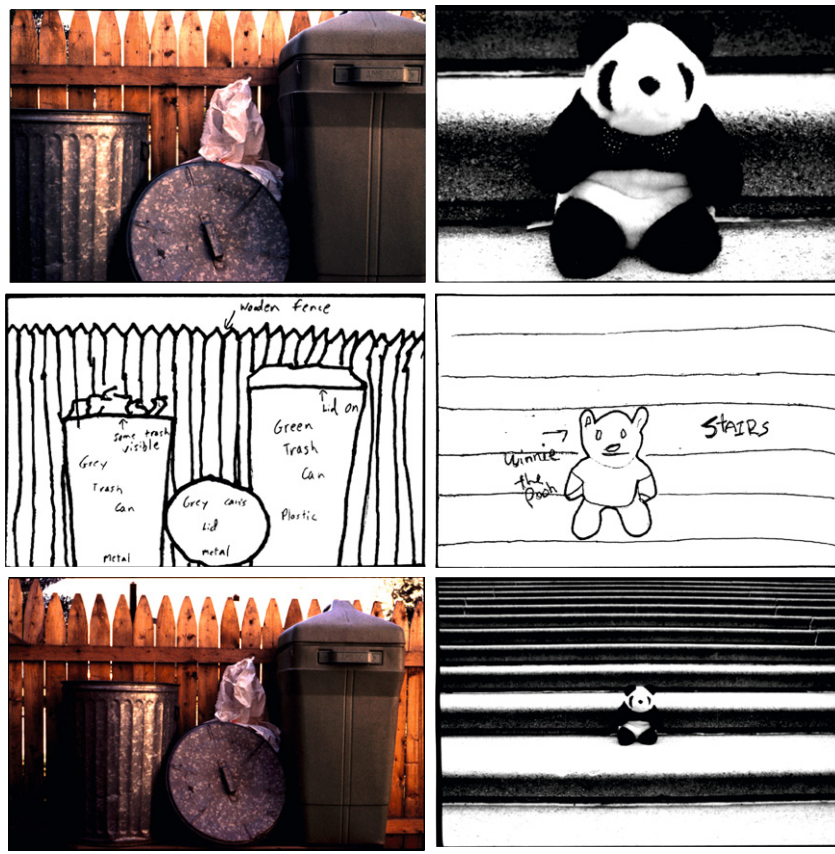

Figure 1 Top row shows close-up views of scenes, middle row shows representative participants' drawings from memory, and the bottom row shows a more wide-angle view of the same scenes. Left column includes part of Figure 1 in Intraub and Richardson (1989) and right column includes part of Figure 1 in Intraub et al. (1996).

memory, we administered our recognition/rating test (in which observers reported if the test picture was the same or showed more or less of the scene on a five-point scale) 2 days after participants had studied 20 pictures for 15 s each. Boundary extension occurred in all of these tests.

Subsequent research revealed that our caution was misplaced—boundary extension was evident in recognition/ratings within minutes of observers viewing 18 pictures for 15 s each (Intraub, Bender, & Mangels, 1992). It was evident in drawings made minutes after viewing seven pictures for 250 ms each (at a rate of 1 every 5 s; see Figure 1, right column). In other experiments, observers' ratings revealed boundary extension when on each trial, three briefly presented pictures (325–333 ms each) were presented in succession followed by a 1-s mask and a test picture that the observer immediately rated (Bertamini et al., 2005; Intraub et al., 1996). The same outcome was observed when the masked interval was decreased to as little as 42 ms (Dickinson & Intraub, 2008): an interval commensurate with a saccade.

In other research, single pictures were presented for 250 ms, masked for 2 s, and then participants adjusted the boundaries of a test picture to recreate the remembered view, and their adjustments resulted in boundary extension (Intraub et al., 2006). Finally, in the most rapid test to date, using the recognition/rating task, Intraub and Dickinson (2008) presented a single picture that was briefly interrupted by a distracting visual noise mask for either 250 ms or 42 ms before reappearing to be rated. On trials where no change at all was made to the picture, observers tended to rate the same view as showing less of the scene than before. A disruption of visual sensory input for less than 1/20th of a second was sufficient for boundary extension to occur in memory for a single picture on each trial.

This poses a vexing problem for current models of scene processing. In general, these models work as follows. Given a brief presentation of a picture (e.g., 250 ms, a ''fixation's worth'') with no subsequent mask, a visual sensory representation will be briefly maintained in the visual sensory register (e.g., Loftus, Johnson, & Shimamura, 1985). If the stimulus presentation is masked, aspects of the visual input will immediately be retained in one or more (depending on the specific model) very short-term memory buffers. Candidate buffers include transsaccadic memory (Irwin, 1991, 1993), visual short-term memory (VSTM; Phillips, 1974), conceptual short-term memory (CSTM, which momentarily stores the general concept of the scene; Potter, 1976, 1999), and ultimately, if attention is maintained, some of this information may be consolidated in long-term memory.

These models are well established in the field of visual cognition and have motivated critical questions about how quickly a picture of a scene can be identified (within 150 ms; Potter, 1976; e.g., Thorpe, Fize, & Marlot, 1996; VanRullen & Thorpe, 2001), what elements in the picture trigger rapid scene categorization (global layout characteristics are sufficient,

e.g., Biederman, 1981; Greene & Oliva, 2009), how much of the scene can be retained once the stimulus is gone (a controversy that persists; Henderson & Hollingworth, 2003; O'Regan & Noë, 2001; Simons & Rensink, 2005), and what poststimulus factors influence memory for a scene (e.g., visual masking vs. conceptual masking; Intraub, 1984; Loftus & Ginn, 1984; Potter, 1976). However, this type of model neither predicts nor provides a good explanation for a rapidly occurring error of commission. To explain boundary extension within this framework, we must either propose the existence of yet another very short-term buffer (e.g., a "scene extrapolation buffer") or add this computational feature to a limited capacity buffer already in the model (e.g., transsaccadic memory)—a solution that would incorporate boundary extension into the model, but in an *ad hoc* manner that does not have explanatory power.

2. Scene Perception as an Act of Spatial Cognition

I would like to present an alternative framework for consideration—a *multisource* model of scene perception (Intraub & Dickinson, 2008) in which scene representation is not conceptualized as a visual representation, even when the stimulus is a picture. At its core, scene representation is an act of spatial cognition. An egocentric spatial framework serves as the "bones" of scene perception, that are then "fleshed out," in the case of a picture by visual representation, amodal perception, and associations to both general and specific contexts. The approach I will describe bridges three subfields of cognition that are generally studied in isolation from one another: visual cognition, spatial cognition, and models of long-term memory (specifically, *source monitoring*; Johnson, Hashtroudi, & Lindsay, 1993; Lindsay, 2008). It also has the benefit of providing a foundation for thinking about scene perception in a way that is not necessarily tied to the visual modality. The model requires a fresh look at some of the definitions and assumptions that have guided research on visual scene perception since the 1970s. I will discuss a new take on some old terminology that I believe can benefit discussion of scene perception in general, but that are required to lay a foundation for explaining the multisource model.

2.1. Definitions: What is a Scene?

Is the photograph in Figure 1 (top of left column) a scene? The expected reply would be a resounding "yes" because scenes are most frequently described in the literature as *views of the world*. For example, in their working

definition of a scene, Henderson and Hollingworth (1999) captured the implicit characterization that is embraced by many:

> In research on high-level scene perception, the concept of scene is typically defined (though often implicitly) as a semantically coherent (and often nameable) view of a real-world environment comprising background elements and multiple discrete objects arranged in a spatially licensed manner. Background elements are taken to be larger-scale, immovable surfaces and structures, such as ground, walls, floors, and mountains, whereas objects are smaller-scale discrete entities that are manipulable (e.g., can be moved) within the scene (p. 244).

The authors themselves were quick to point out that this working definition has shortcomings; in particular, they raised the problem of scope, for example, at what point is a picture too close-up to be considered a view of a scene (e.g., the content of a drawer) or too encompassing (an aerial view)? The problem of scope, they pointed out is typically avoided in studies "...by using views of environments scaled to a human size. So an encompassing view of a kitchen ... would be considered a good scene..." (p. 244). This working definition has lead to other problems as well, notably disagreements about whether or not a particular set of pictures can "count" as being a set of *scenes*.

For example, if multiple objects are required before a picture can be a scene, then the photograph in Figure 1 (top of right column), which contains a single object against a background, is not a scene. A tricky claim, because it was in fact a view of the world snapped by a photographer. Requirements such as "semantic coherence" and "spatially licensed" organizations and the expectation of "encompassing views" have also raised problems. For example, Zelinsky and Loschky (2005) needed to justify their choice of photograph of toys in a crib for a visual search experiment that they argued was relevant to search in the world. This was because nothing constrained the placement of the toys (in the way, for example, the placement of large appliances are constrained in a kitchen). However, again, the close-up view of the crib is a plausible view that a parent or baby might frequently see. Intraub (2004) needed to justify the use of objects on desktops and on portions of the floor in a 3D boundary extension experiment, because these were not "encompassing" views. However, they are normal views during day-to-day perception. Why should these stimuli be deemed unsuitable for studying aspects of scene perception and memory?

I believe that many of these arguments arise because the definition of a scene throughout the field frequently focuses on acceptable characteristics of the displays that we use in our research, rather than on characteristics of the 3D world. Our view of a kitchen is sometimes an encompassing view, but not always. In a small kitchen, a view of the stove may disallow a view of the refrigerator because it is on the opposite wall, behind the observer. In the world, viewers are embedded within the scenes they observe. We do stand

by a desk and look down at the desktop, or kneel on the floor to pick up a child's scattered toys.

The ultimate goal of scene perception research, after all, is to understand how we perceive our surroundings, not how we understand pictures. This goal is clearly laid out in most papers, but then later on, the same term, "scene" that is used to refer to the world, is also used to refer to the pictures that serve as stimuli. What I will argue is that a photograph can never be a scene nor can it show an entire scene; any more than an observer can see his or her surroundings *all at once* (e.g., Hochberg, 1986; Intraub, 2007). Eye movements, head movements, and body movements are necessary to explore scenes in the world; we are embedded within the scenes with which we interact. I suggest that the term *scene* be used only to refer to surrounding environments in the world. The pictures we use in our research are not in themselves scenes. Throughout this chapter I will refer to them as *proxy views* to remind the reader (and myself) that the pictures we use in our research are only 2D surrogates for actual views of the scenes that surround us in the world. I suggest that *scene representation* be reserved for referring to the mental representation of the surrounding 3D world elicited by a view. If we ask people to remember the proxy view we showed, consider that we are only asking them about one part of their scene representation—their memory for the pictured view.

These definitions will allow us to clearly separate four distinct concepts in discussing research on scene perception and memory: (a) the surrounding 3D scene in the world where the picture was taken, (b) the 2D proxy for a view of that world (a photograph, computer-generated image, or line-drawing), (c) the mental representation of the surrounding 3D scene elicited by the view, and (d) representation of the details originally contained in the proxy view [e.g., textures, colors, objects (identity, orientation, and so forth)].

By accepting the idea that a scene refers to the world surrounding an observer, then any view (a wide-angle view of a kitchen or a close-up view of a desk top) is a view of a scene. A view can contain a single object, a pile of objects bearing a scrambled relation (e.g., a picture of a junk yard), and the inside of a draw. Carmela Gottesman and I have suggested that a more fundamental characterization of what a picture must include to be understood as a view of a scene is the depiction of a location that is cropped by the picture's edges (Gottesman & Intraub, 2002; Intraub, Gottesman, & Bills, 1998). Thus, even an aerial photograph (as in Zelinsky& Schmidt, 2009) is a view: it is a view from a satellite that continues beyond the edges of the given photograph.

2.2. An Illustrative Anecdote

To illustrate the concept of a multisource scene representation, I will provide the following anecdote. Recently, while preparing a talk, I looked closely at the picture at the top of the left column in Figure 1 and was

shocked to realize that I may have misunderstood it for the past 20 years! I had always perceived the picture to be a view of trash cans awaiting trash pick-up at the curb in a suburban neighborhood. I thought that the tripod and camera were positioned in the street, with a nonspecific suburban house behind the camera. In fact, I had acted on this interpretation, requesting that subsequent student photographers refrain from setting tripods in the street for safety reasons. What I had now focused upon were the crossbeams of the fence, noticing that they were the type that usually (but not always) signify the "inside" of a fence. The picture may not have been taken on the street at all!

My curiosity piqued, I decided to ask a colleague if based on his memory for the picture, he had a sense of what was behind the camera. Without hesitating he immediately replied, "The owner's house." When queried, he said that he had *always thought of it* as trash cans in the owners' backyard, opposite their house. Was the picture taken in a backyard or out on the street? I asked my graduate student what she thought was behind the camera in the hope of breaking the tie. Without hesitating she replied, "The other side of the alley." I was nonplussed—what alley? The ensuing conversation enlightened us both as to the trash storage methods in different suburban areas of our country. In the southwest where she had lived all her life, trash cans were kept in fenced alleyways between suburban homes—something that neither I nor my colleague had experienced in the eastern and Midwestern suburbs with which we were familiar.

We all saw the same proxy view, we all remembered the key content of the proxy (the trash cans, the lid, and the wooden fence), but in addition, we all adamantly claimed to have had a *scene representation* that captured our understanding of the view and the space in which the camera had been embedded (i.e., "I *always* thought of it that way"). The representation went beyond the picture and beyond the region that is included when boundary extension takes place. It included the world behind the viewer. Yet, it is important to note that none of us believed that we had seen what was behind the camera.[1]

2.3. A Multisource Model of Scene Representation

As mentioned earlier, traditional approaches to scene processing have focused on memory for the content of the proxy view. The representation is thus based on a single source—the visual input. Intraub and Dickinson (2008) proposed that instead scene representation might be best conceptualized as a multisource representation, even when the observer is simply viewing a picture. The "starting point" in this approach is not the visual sensory input as in the traditional approach to scene perception, but is an underlying spatial structure (frame of reference) that the human observer

[1] I thank James E. Hoffman and Kristin O. Michod for sharing these descriptions.

brings to the event (whether the experience involves exploring the 3D world, watching a movie, or looking at a picture). Researchers have explored many different kinds of spatial reference frames and different terminologies are used to describe them. Three general categories and typical terminologies include *egocentric* (e.g., right now, the desk is front of me), *allocentric* (e.g., the armchair in on the wall opposite the desk, adjacent to the door), and *geographic* (e.g., the room is in a house, at the southwestern corner of Pennsylvania; see Allen, 2004, for discussions of different ways of conceptualizing frames of reference). Thus, in this view, at its core, scene perception is an act of *spatial cognition*.

In the case of a single novel view, the most prominent framework would be an egocentric frame of reference that includes the observer's sense of "in front of me," "on my left," "on my right," "above me," "below me," and "behind me" (for a discussion of imagined space, see Bryant, Tversky, & Franklin, 1992; Franklin & Tversky, 1990; Tversky, 2009). Although I will focus on the egocentric reference frame here, it is expected that other frames of reference can also be activated by a proxy view (e.g., a geographic reference frame, as when asked the location of a familiar store in a picture; Epstein & Higgins, 2007).

The key assumption is that a proxy view will trigger a number of mental activities: (a) visual processing of the proxy view, (b) amodal perception of the objects (Kanizsa, 1979) and surfaces (Nakayama, He, & Shimojo, 1995; Yin, Kellman, & Shipley, 2000) just beyond the boundaries of the view, (c) categorization of the view (e.g., basic-level categories of natural landscapes such as desert, field, forest, lake, mountain, ocean, and river; Greene & Oliva, 2009) and contextual associations elicited by objects in the view (e.g., Bar, 2004). These multiple sources of input are organized within the egocentric spatial structure surrounding the viewpoint taken by the observer. In the case of a photograph, the observer takes the viewpoint of the camera (as is also true in viewing films; Hochberg, 1986; Intraub, 2007).

Just as the visual field is graded, decreasing in resolution from the fovea outward (see O'Regan, 1992), scene representation is also graded, extending well beyond the visual information in the proxy view. For example, consider a briefly presented picture with the observer fixating the center. The best visual resolution would be at the point of fixation, shading off toward the boundaries of the picture. Where the visual sensory input abruptly ends at the picture's boundaries, perception does not end. Amodal perception allows the viewer to *perceive* the scene beyond the edges of the view by completing any objects cropped by the picture's boundaries (e.g., Kanizsa, 1979) and by continuing cropped surfaces (Fantoni, Hilger, Gerbino, & Kellman, 2008; Yin et al., 2000). Although this is amodal perception (not perception of the visual input), it is nonetheless crucial for comprehending the view. Without it, observers would interpret the proxy view in Figure 1 (top of column 1) as a view of broken trash cans

(i.e., chopped in half) and a fragment of a broken fence (with the top of the pickets chopped off). But that is not what we perceive. The visual input at the picture's edges tightly constrains amodal perception just beyond the boundaries of the view (cf. Nakayama et al., 1995). World knowledge would support these constraints as well (one knows the expected shape of a trash can). Indeed, all of the participants' drawings of this proxy view in Intraub and Richardson (1989) depicted *unbroken* (whole) trash cans and the *continuation* of the fence at the top and side boundaries, as well as the lid at the bottom (as shown in Figure 1).

Constraints become less specific for regions that are farther away from the visual information (e.g., the "nonspecific suburban house" behind the camera mentioned earlier, or in the case of a nonspecific location, such as a close-up of a cup on a table, merely "a wall in the room," behind the camera, without a particular type of room being specified). Thus, many aspects of scene representation that were not visually present in the proxy view will be shared across observers: the continuation of the view just beyond the boundaries (as in boundary extension), the categorization of the view as an outdoor scene (that must have a sky above), and at least for viewers familiar with these types of trash cans and fences in the United States, the interpretation of the view as a suburban scene. However, individual experiences will cause divergences (e.g., as when different observers interpreted the trash cans and fence as being in an alley, in front of the house, or in the back yard).

When I suggest that scene perception is an act of spatial cognition I mean that all of these sources of information are organized within an egocentric spatial framework, resulting in a scene representation that is laid out in terms of the space around the viewer. While the proxy view is *visible*, viewers have no difficulty discriminating the currently present *visual sensory input* from the *amodally perceived* continuation of objects and surfaces beyond the view-boundaries. Put simply, one can simultaneously perceive that these are normal, intact trash cans, yet report that the picture only permits us to see a part of each one. Given the ease of discriminating these sources of information when the picture is present, why then do observers falsely remember having *seen* beyond the edges of the view when the visible stimulus is interrupted for less than 1/20th of a second before reappearing at test (i.e., *boundary extension*; Intraub & Dickinson, 2008)?

2.4. Boundary Extension as a Source Monitoring Error

According to the multisource model, once the sensory input is gone, even for a moment, the observer's only recourse is to rely upon memory of the experience. That is, *memory* for information that had originally been visually perceived (which itself varies in terms of resolution), *memory* for information that had originally been generated through amodal perception, and *memory* for the contextual layout that had been elicited. Although we ask the

observer to remember the proxy view and compare it with the test item, it is not simply memory for the proxy view that the observer has. Instead, the observer has in mind the representation of a scene and must now determine which part of that scene representation had been contained within the boundaries of the proxy view—that is, had a visual *source*. Thus, although unintended, the boundary memory task may in fact be a *source monitoring* task (Johnson, 2006; Johnson et al., 1993; Lindsay, 2008), more specifically a *reality monitoring* task (Johnson & Raye, 1981), because we are asking people to decide where the externally presented information ended and the amodally generated information began.

The fundamental insight of Johnson and her colleagues is that the source of a memory is not stored in the form of a tag that specifies where the information came from. In most cases, determining the source of a memory is an *attribution* that is based upon the type and quality of details (perceptual, contextual, semantic, or emotional) that characterize that representation (Johnson, 2006; Johnson et al., 1993; Lindsay, 2008). Different sources are associated with different profiles of characteristics, and source monitoring is a process by which the individual determines which profile the memory best fits. Most of the time, memory may fall squarely within a category (e.g., ''I saw that with my own eyes''). But sometimes, it may not. So, for example, if memory for a dream includes characteristics than are unusual for dreams (very high levels of perceptual detail, a well-integrated context, strong emotional response, and clearly defined co-temporal events) but are hallmarks of perceptually based experiences, one might mistakenly attribute the experience to perception. This also explains mundane mental puzzles such as a common rumination that plagues travelers after leaving for a trip: did I actually turn off the stove before leaving, or did I just think about turning it off? Source monitoring has been able to provide an account of many well-known long-term memory errors (see Lindsay, 2008 for a review). What we propose is that the same monitoring process can explain boundary extension.

In this view, the boundary extended region people erroneously remember having seen is not computed in a very short-term buffer after the stimulus is masked. The continuation of the view was always part of the scene representation (i.e., the trash cans were always perceived as whole, and the background was always perceived as continuing beyond the edges of the view). Although readily distinguishable while the visual sensory input is available, distinguishing the difference between *memory* for the visual information at the periphery of the picture and *memory* for the amodally-generated information just beyond it is much more difficult. The individual must decide at what point the image is just not detailed enough to have been visual. This results in a source monitoring error, in which memory for amodally perceived information that had been tightly constrained by the visual information (and context) is now attributed to having been seen—that is, boundary extension. Although boundary extension is an error with

respect to the proxy view, it is typically a good prediction of upcoming layout just beyond the view and as such may assist in the integration of views (Intraub, 1997, 2002).

2.5. Effects of Divided Attention and Stimulus Duration on Boundary Extension

If boundary extension is a source monitoring error, then we should be able to draw on the source monitoring model to predict ways of influencing the size of the boundary extension error. How might this be accomplished? Once again, while the visual information is present, there is a very sharp and discontinuous delineation between the current visual information and the amodal continuation of the scene. After a brief interruption, the sharpness decreases ("flattens out") in the remembered representation because visual memory is not a photographic representation of the sensory input. Thus, memory for details in the periphery of the picture and memory for the highly constrained amodal continuation of those details will share many characteristics causing a source monitoring error. Perhaps, if we were to flatten the gradient between the two further, the threshold for deciding where the visual information ended would be lowered, and the observer would accept a slightly greater swath of amodally perceived space as having been seen before.

The outcome of a series of experiments that tested the effect of divided attention on boundary extension is consistent with this interpretation. Instead of divided attention resulting in an increase in random errors when rating the remembered boundaries of a proxy view, when visual attention was divided in a dual-task situation, boundary extension increased, that is, more "nonvisually derived" information was attributed to vision than when attention was not divided. In the initial experiments, Intraub, Daniels, Horowitz, and Wolfe (2008) presented participants with close-up or wider-angle versions of simple scenes, similar to those used in previous boundary extension experiments. However, superimposed on each picture were randomly positioned block 2s and 5s (as shown in Figure 2). Visual

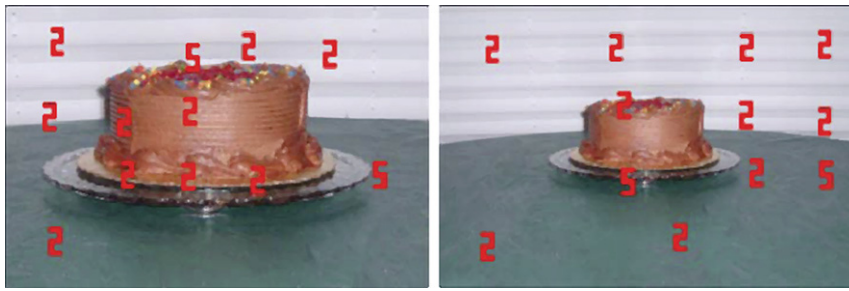

Figure 2 Example of a pair of close-up and wide-angle views with superimposed 2's and 5's. From Intraub et al. (2008).

attention was manipulated with a search task in which observers had to report the number of 5's (there could be zero, one, or two on any trial). This is a very difficult search task that was made all the more challenging by limiting the display time to only 750 ms on each trial and presenting the numerals on a photograph. There were three independent conditions: memory only, dual task (giving the search task priority), or search only. In the memory-only and dual-task conditions, after a masked interval, a test picture appeared and participants rated it on the standard five-point boundary recognition/rating scale (as ''same,'' ''more close-up,'' or ''more wide-angle'' than before).

Although the search task was extremely difficult, participants performed above chance in both the search-only condition and the dual-task condition. Critically, search performance was the same across those conditions, demonstrating that dual-task participants had indeed given priority to the search task. When boundary ratings were compared between dual-task and memory-only conditions, both yielded significant boundary extension; however, boundary extension was greater when attention was divided. In both conditions on trials on which the stimulus and test pictures were the same close-up view, they rated the test picture as looking too close-up (indicating that they remembered the original with extended boundaries). When the stimulus and test pictures were different, the typical distractor asymmetry indicative of boundary extension occurred; the stimulus and test pictures were rated as more similar when the close-up was the stimulus than *vice versa*. This asymmetry signifies boundary extension because when the closer view is the stimulus, boundary extension in memory would cause a wider view at test to be a fairly good match, whereas boundary extension would have the opposite effect were the wider angle view presented first. Clearly, dividing attention did not introduce random error, but instead (in terms of the multisource framework) increased the acceptance of amodally generated information as having been seen before.

Results could not be attributed to observers in the memory-only condition capitalizing on their less demanding task to develop strategies for ''beating'' the expected rating task (e.g., verbalizing, ''the shoe is 0.5 cm from the right edge'') because the same results were replicated under conditions in which instead of being tested after each trial, the memory test was deferred until the end of the experiment and all participants were naïve as to the nature of the test until it was administered. Again, search was above chance and did not differ between the baseline-search condition and dual-task condition boundary extension was greater in the dual-task than memory-only condition, and the rating asymmetry in response to distractors was obtained.

Additional tests of the source monitoring hypothesis will have to be conducted. Some tentative support comes from a comparison of boundary extension for the same pictures shown at different stimulus durations.

If we accept that the overlap in similarity between visually-generated and amodally-generated memory might be greater if visual detail were reduced relative to the highly constrained amodal information, then reducing stimulus duration might be another way to "flatten" the difference between information from the visual source and that derived from amodal perception. Intraub et al. (1996) reported greater boundary extension for 250-ms pictures than for 4.5-s pictures when they were shown at the same rate (1 every 5 s). Similar to the divided attention experiment, rather than introducing more random error into the ratings, the briefer stimulus duration resulted in greater boundary extension. However, this was only a single test with a very small stimulus set (seven stimuli). In ongoing research, Christopher Dickinson and I found in one experiment that boundary extension increased as stimulus duration decreased from 500 to 250 to 100 ms (with rate of presentation held constant). This occurred for multi-object scenes. However, in other research, using tight close-ups of single objects, no difference was obtained as a function of stimulus duration. In this case, the amount of boundary extension was very great in all conditions, so that it may have swamped any detectable differences at the durations tested, but the results at this point are unclear. Future research will explore this further and also explore other means of increasing or decreasing the difference between memory for the internally generally amodal perception and memory for the visually presented information.

3. Multisource Scene Representation: Behavioral and Neuroimaging Picture Studies

In the visual cognition literature, pictorial stimuli have often been referred to as if they were endpoints on a scale of simple to complex visual stimuli. For example, in early research on scene perception, Potter (1976) described the photographs in her study as "complex, meaningful visual events." However, as research has progressed, we find that views of scenes may instead be a distinctive class of stimuli, that differ in important ways from other types of visual stimuli.

3.1. Denoting a Location: The Importance of View-Boundaries

Beginning with behavioral research, boundary extension is not elicited by all picture boundaries. If a display does not depict a location—a view of an otherwise continuous scene—boundary extension does not occur. This is not to say that observers' memories are error free, but that instead of reflecting boundary extension, they tend to be bidirectional indicating an

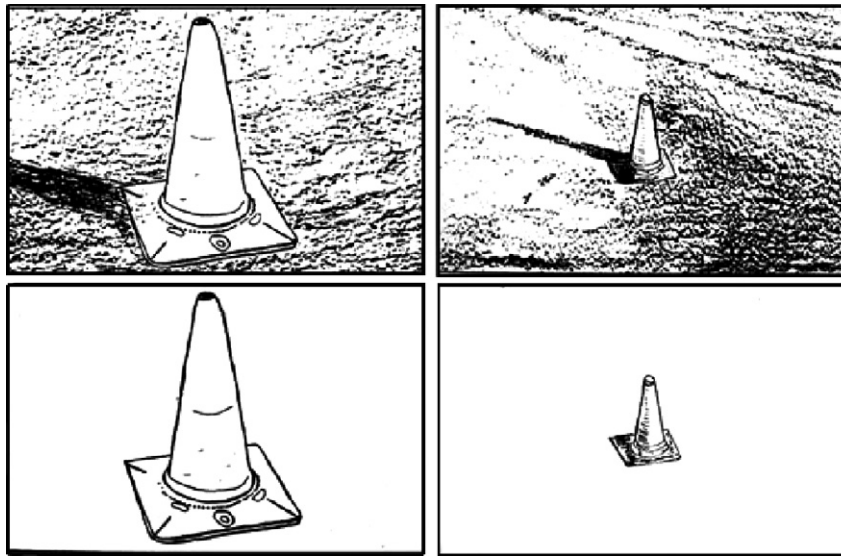

Figure 3 Example of close-up and wider angle views of an object on a blank background versus a meaningful background presented to participants in Intraub et al. (1998).

averaging effect—regression to the mean object size (Intraub et al., 1998; also see Gottesman & Intraub, 2002; Legault & Standing, 1992). Figure 3 shows an example of pictures in which the same sized objects are presented both with and without a location specified (i.e., a blank background or a background other observers had described as an asphalt road; Intraub et al., 1998). A pattern of errors consistent with boundary extension occurred for pictures containing backgrounds (i.e., that showed a partial view of a continuous scene). However, the error pattern was different when only a blank was present in the background—in this case, regression to the mean object size occurred instead (i.e., large objects remembered as smaller, and small objects remembered as larger).

This suggests that depiction of a background is important for a multisource scene representation to be elicited. However, we found that if we recruited participants' scene knowledge, and required them to imagine the specified background (e.g., an asphalt road with a shadow of the cone) while viewing pictures of objects without backgrounds, then boundary extension occurred. In fact, the ratings were indistinguishable from those obtained when the background was visually presented. We demonstrated that this was not an artifact of requiring an imagination task, because when another group of participants was required to imagine colors on the outline objects with blank backgrounds (and no mention of a scene was made), then the error pattern shifted back to one that reflected regression to the mean object size.

In the same vein, Gottesman and Intraub (2002) demonstrated that either boundary extension or regression to the mean object size would be observed depending on whether people were led to interpret a blank background as being the location on which an object was photographed (i.e., being part of a scene), or an unrelated background. The latter was achieved placing a cutout of a photographed object on a white background in front of the observer. Ultimately, the two conditions were the same, an object on a white background. When that background was understood as being unrelated to the picture, a bidirectional averaging error occurred (big objects were remembered as smaller, and small objects as larger), but when it was understood as depicting the location at which the picture was taken, boundary extension occurred. In terms of the model, a multisource scene representation had been elicited. Interpreting the edge of the picture's, white background as being a *view-boundary* appeared to be a critical factor in eliciting boundary extension.

The distinctive nature of view-boundaries has been illustrated in other research in which different "types" of surrounding boundaries were compared. In Figure 4 are two of several proxy views that were presented to participants in Gottesman and Intraub (2003). Both pictures yielded boundary extension beyond their edges. Both also include a surrounding border *within* the picture. In the picture of the sandal and towel on the grass, the edges of the towel surround the sandal; these edges parallel those of the picture's view-boundaries, but are not themselves view-boundaries—they are the edges of an object (the towel), not the edges of a view. In the picture of the desktop, however, among the objects on the desk is a framed photograph and that photograph has its own view-boundaries. Participants did not remember seeing a greater expanse of the towel around the sandal, but did remember having seen a greater expanse of the background around the fork (in the picture within the picture). Thus, a boundary

Figure 4 The edges of both pictures provide a view-boundary: inside the picture on the left is an object boundary (the edges of the towel that surround the sandal) and inside the picture on the right is another view-boundary (the edges of the picture on the desk, i.e., the picture within a picture). Based upon two figures in Gottesman and Intraub (2003).

inside a picture appears to elicit boundary extension only when it is a view-boundary. Similarly, DiCola and Intraub (2004) demonstrated that when an object was occluded by a view-boundary, participants remembered having seen more of the object than was visible. However, in contrast, when the same object was occluded by another object within the scene (identical amount of occlusion), this unidirectional error did not occur.

3.2. Boundary Extension and Scene-Selective Regions of the Brain

There are parallel observations in neuroimaging studies that suggest special properties of views of scenes. Similar to the presence or absence of boundary extension as a function of whether the background depicts a view of a scene or is blank, the parahippocampal place area (PPA) was reported to respond most strongly to views of locations in space (e.g., a room with objects, or an empty corner of a room with no objects), but respond poorly to objects that were not depicted in a location. Similar to the behavioral work showing that imagining a background when looking a picture with a blank background results in boundary extension (Intraub et al., 1998), PPA responded strongly when participants were required to imagine a location (O'Craven & Kanwisher, 2000).

Whereas PPA is thought to respond most strongly to the local spatial layout in a specific view, the retrosplenial cortex (RSC) is thought to be involved in integrating a local view within a broader spatial context—perhaps being related to navigation and recognition of places (Epstein & Higgins, 2007). To determine if these scene-selective areas would respond to boundary extension in memory, Park, Intraub, Yi, Widders, and Chun (2007) presented observers with series of photographs in which there were repetitions. In the critical conditions, the repetition was not an identical view; a close-up view would later be followed by a wider view of the same scene or *vice versa*. Observers were simply instructed to remember the photographs. The use of repetition was a critical feature of the design; reduction in the neural response in a predefined area of the brain the second time a stimulus is presented is thought to indicate that the brain area has treated the two stimulus events as being the same (a habituation effect; Turk-Browne, Scholl, & Chun, 2008). Put simply, novel items should elicit greater activity than repeated items.

The rationale was that if PPA and RSC are not responsive to boundary extension in memory, then repetition of the same scene (whether a slightly wider view was shown first or second) should result in similar reductions in the neural response. However, if these brain areas respond to boundary extension (i.e., if the extended region is accepted as having been seen before) then, following the asymmetrical response pattern described earlier, a closer picture followed by a wider picture should result in greater

attenuation than a wider picture followed by a closer picture. Why? Because if the closer picture is first and is remembered with extended boundaries, then when the wider view is presented it should look somewhat similar to what the observer remembers (as in "seen that before"), whereas if the wider view is first, the close-up would appear to be a very different view, with any boundary extension exaggerating the difference (as in "that's new").

As shown in Figure 5, the PPA and the RSC responses yielded an asymmetry. When a closer view was followed later by a wider view, the neural response to the wider view decreased, but when the wider view was followed by the closer view, the neural response to the second view was just as strong as it was to the first (indicating that the region responded to this picture as if it were new). This asymmetry was not observed in the lateral occipital cortex (LOC), which is associated with object recognition. In this area, the size of the object is not expected to matter, just its identity, so whereas PPA and RSC showed the asymmetry, LOC clearly did not; the same habituation of the neural response occurred irrespective of which view was presented first (as in "seen that before"). These results suggest that both PPA and RSC are sensitive to boundary extension.

However, a subtle difference between the two conditions supported the distinction described earlier between PPA and RSC. Although the PPA showed some sensitivity to boundary extension in these critical conditions, it also responded as if it was retaining some of the specific layout information from the local proxy view because neural attenuation occurred in the two identity conditions (i.e., a close-up followed by the same close-up or a wide-angle view followed by the same wide-angle view). In other words, repetition of the close-up was recognized as a repetition of the same view. So the results were mixed for PPA. However, the RSC responded to the identity conditions differently, showing no habituation; there was no attenuation of the level of response when the same pictures were repeated. This provides converging evidence for the idea that RSC responds to the integration of an individual view within a larger scenic structure. The lack of attenuation in the identity conditions suggests that the first presentation had been remembered within a larger scene context, and thus its repetition was responded to as novel. Results suggest that both scene-specific areas responded to boundary extension to some degree, but that the RSC was more attuned to placement of the specific view within a more expansive framework.

4. Multisource Scene Representation: Exploring Peripersonal Space

Pictures do not surround the observer, usually subtend a relatively small visual angle (as compared with the scope of the visual field), are 2D representations, and in many ways differ dramatically from the experience of

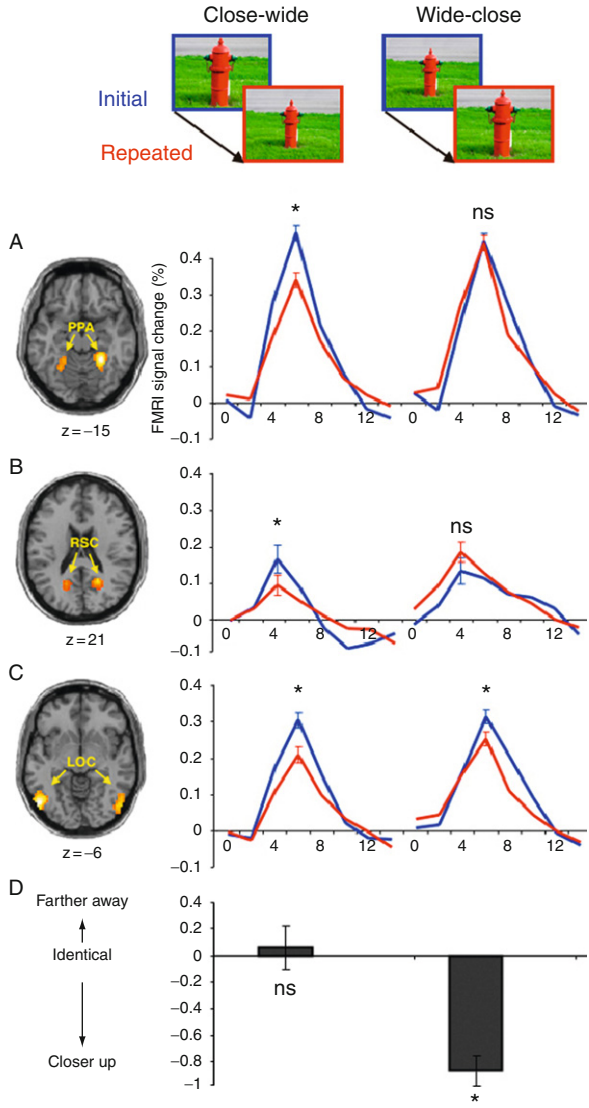

Figure 5 Boundary extension in the PPA and RSC but not in the LOC: a representative participant's PPA, RSC, and LOC. ROIs are shown on a Talairach-transformed brain. Examples of the close-wide and wide-close viewing conditions are presented in the top row. Hemodynamic responses for close-wide and wide-close conditions are shown for each ROI. Error bars indicate standard error of mean (±S.E.M.). Bottom row shows the same asymmetry in behavioral responses of these participants in a test outside the scanner. Based on Figure 2 in Park et al. (2007).

viewing the world. Still they have proved to be valuable tools that can allow us to learn about scene perception. It is generally assumed that the similarities between viewing pictures and viewing the world outweigh the differences. I have always embraced this position, although as I began to think about boundary extension in more spatial (than visual) terms, I became concerned about the validity of this assumption. This is because in real space when looking at an occluded view (e.g., through a window), rich spatial information derived from stereopsis, motion parallax, and the relation of the edges of that view to one's body could serve to constrain scene representation, and prevent an error like boundary extension from occurring. If boundary extension is fundamental to scene perception, then if we were to set up views in the real world that bear similarity to those presented in pictorial form, we would expect people remember seeing beyond the edges of the view in the world as well. If it is picture memory error, then people may be able to remember the location of the edges of a view through a window that is directly in front of them.

If boundary extension did occur in real space, then this would also provide an opportunity to test another key assumption of the multisource model. If at its core, scene perception begins with a surrounding spatial framework that is then filled in by various sources of information, then we should be able to see some similarities between scene representations that are initiated through vision and scene representations initiated through touch (more specifically touch and movement, i.e., haptic exploration). If a person were to feel a space within the boundaries of a window-like opening, would they experience boundary extension, that is, would they remember having *felt* beyond the boundaries of the view?

Unlike vision, which is a distal sense, with a very small foveal region and large low-acuity periphery, touch is a contact sense with multiple high acutely regions (i.e., the five fingertips, which can be thought of as five "foveae" on each hand) and a relatively small low-acuity periphery (e.g., the size of a hand). The span of the hands is much smaller than the span of the eyes. It is possible that these differences would make it more likely that people could correctly retain the expanse of the "view." These differences may allow for memory beyond the view in the case of vision but not in the case of haptic exploration. In fact, the notion "touch teaches vision" has a long history in psychology—including the sense that haptic input can serve to test the reliability of certain types of visual cues (see Atkins, Fiser, & Jacobs, 2001).

However, in spite of the many differences between vision and haptics, the brain is presented with a common challenge: in the case of both modalities, a coherent continuous representation of the world must be established based upon successive, discrete, sensory inputs. Whether we are viewing a room or haptically exploring a room in the darkness, we can never explore it all at once—the environment must be sampled a part at a time.

This point is one that William James drew upon in his seminal argument against the theory that blind individuals must represent the spatial world in a radically different way than sighted individuals. He pointed out that the apparent piecemeal nature of haptic input is no different than the ''innumerable stoppings and startings of the eyeballs'' during visual perception, and that these two examples of successive inputs are likely integrated in similar ways (James, 1890).

4.1. Haptic Exploration: Sighted Observers and a Deaf and Blind Observer

To begin to explore these questions, Intraub (2004) set up the following conditions. Six small regions in two adjacent rooms were set up on table tops and on the floor, with a window-like apparatus around each. Semantically related objects were placed in reasonable relation to one another within each, as shown in Figure 6. The ''windows'' used in the visual condition were attached to an expanse of cloth to block participants' view of the background outside the window, and the window frame was very flat so that it would not occlude participants' view of the surface within the window. In addition the window frame was placed directly on the background surface so that no matter how the viewer shifted his or her position, we would know for certain that he or she could not see beyond the edges of the view (which is not the case for a typical window which is positioned between the viewer and the viewed surface). In the haptic condition the window frame provided the same sized ''view,'' but was made of wood that was high enough to prevent people from accidently feeling outside the stimulus area when they explored the regions while blindfolded. Examples of the two types of windows are shown in Figure 7.

Similar to the proxy views (close-up pictures) used in boundary extension research, observers were positioned so that they were very close to the stimulus areas (as shown in Figure 7). Thus, these stimuli allowed natural views in real space that were similar to views already tested in other research (e.g., Gottesman & Intraub, 2003). Close viewing also provided a conservative test of whether or not boundary extension would occur when *viewing* real spaces because in peripersonal space (where viewed objects are close enough to grasp) one would expect distance and area judgments to be more accurate than if one had studied a distant view through a typical window (see Previc, 1998, for a review of theories regarding perception of near versus far space).

All observers were blindfolded and escorted to each stimulus region so that they would experience it only from the experimenter's designated viewpoint. In the vision condition, the blindfolds were removed for 30 s while observers studied the view. In the haptic condition, observers' hands were placed at the center of the region and they were instructed to feel

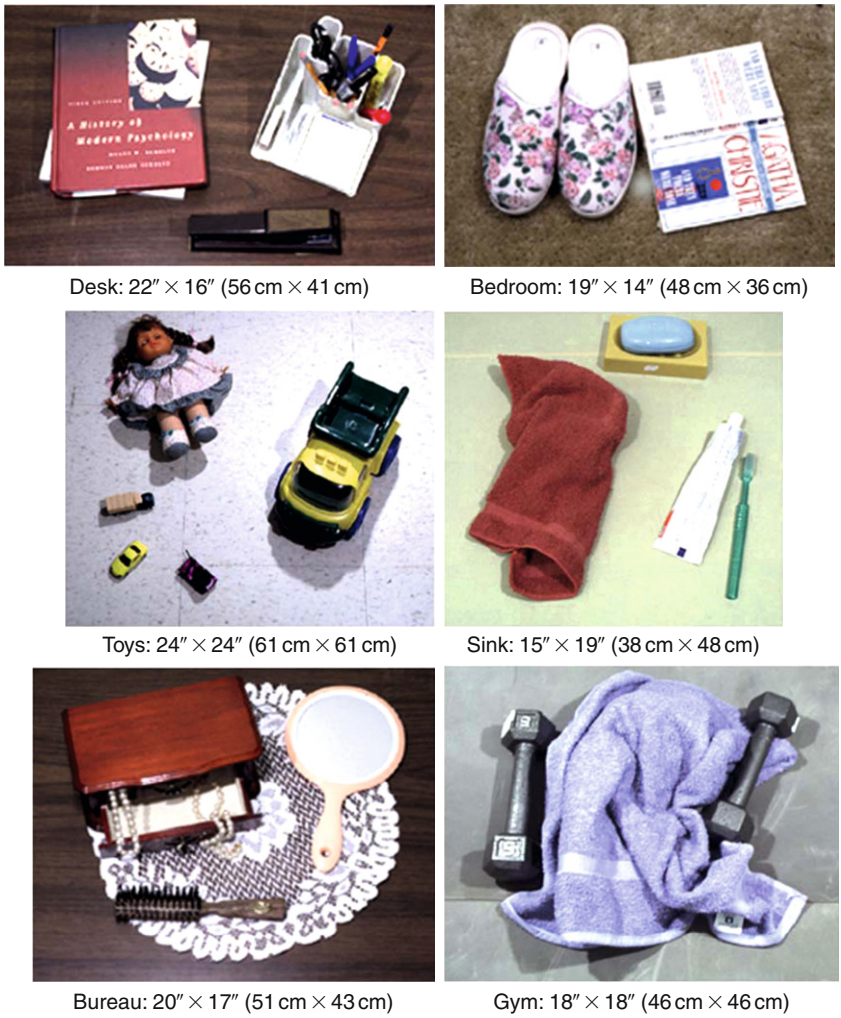

Figure 6 Stimulus regions and their dimensions (photographs were cropped to approximate the view through the "window"). Based on Figure 2 in Intraub (2004).

everything up to but not outside the wooden boundaries during the 30-s inspection interval. In both cases, they were told to remember the areas in as much detail as possible. In all conditions, participants named the objects and described a title for the view (so that we could check that identification and interpretation of the regions were the same across conditions—they were). With the blindfold in place, after having studied each of the six regions, participants were escorted to a waiting area, while the windows were removed by the experimenters.

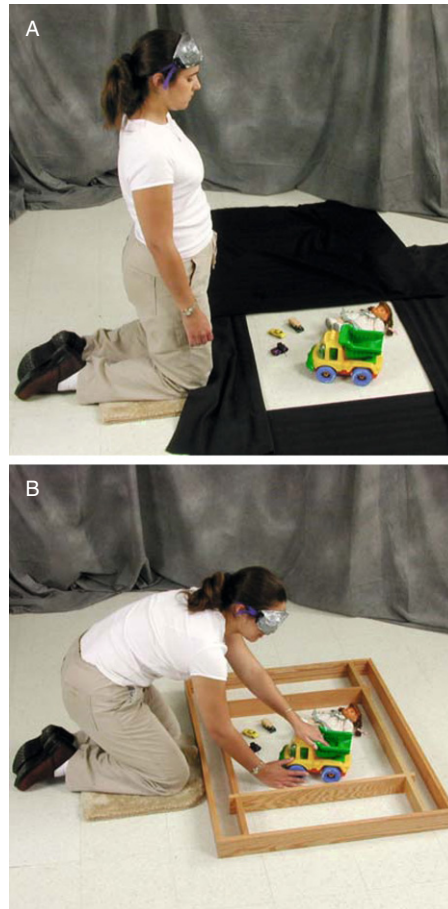

Figure 7 Visual exploration (A) and haptic exploration (B) of the "toys" scene (all borders were removed prior to test). Based on Figure 1 in Intraub (2004).

Upon returning to the regions, participants, using the same modality as during study, were asked to use a fingertip to show where each boundary had been located. Experimenters set the borders down at those locations. Participants were then allowed to make any adjustments necessary to the four boundaries to make the region the same as before. In spite of their proximity to the windows and to the graspable objects, vision participants placed the boundaries out farther than they had been placed originally. The mean area remembered for each scene in each condition is shown in Figure 8. Vision participants increased the area of these views such that the mean area increase across scenes was 53% (which reflected boundary extension in both the length and width of the window). Haptic exploration

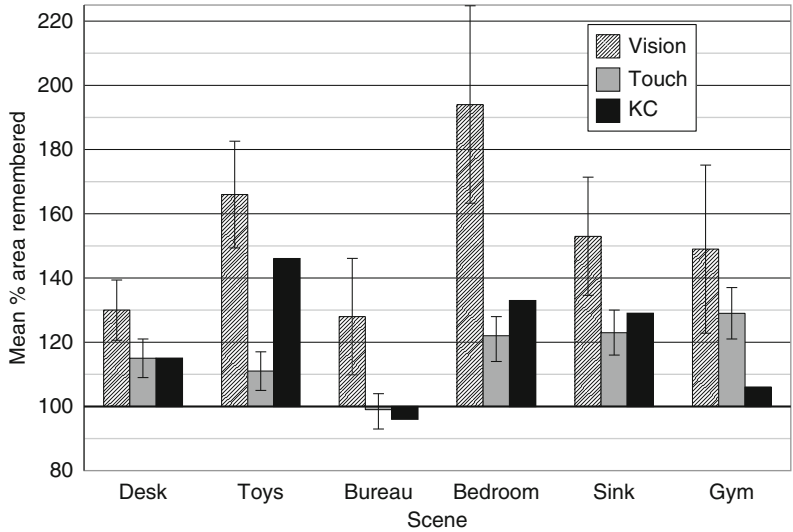

Figure 8 Mean percentage of the area of each region remembered by sighted participants in the visual and haptic conditions and the percentage of each region remembered by KC, who has been deaf and blind since early life. Error bars show the 0.95 confidence interval around each mean. (Boundary extension occurs when the mean remembered area is significantly greater than 100%, i.e., when 100% is *not* included in the confidence interval.) From Intraub (2004).

without vision occurred in five of the six scenes for the blindfolded-sighted participants (the scene that did not yield boundary extension was remembered in a distorted manner because of an alignment illusion in the blindfolded condition; see Intraub, 2004 for details). Although robust, and sizeable, the amount of boundary extension was clearly less than in the visual condition: on average, they increased the area by 17%.

Did boundary extension truly reflect haptic exploration, or because we were not testing ''haptic experts'' but people who were momentarily blindfolded, might this smaller sized boundary extension be the result of participants using visual imagination to support their exploration (as discussed earlier, there is evidence that visual imagination can induce boundary extension; Intraub et al., 1998). To address this issue, there was a third condition in the experiment that included a single participant. KC was a 25-year-old student who had been both deaf and blind since early life. Her natural mode of exploring the world is through haptic exploration, and as shown in Figure 8, her performance was very similar to the blindfolded-sighted observers. As shown, sometimes her area increase was greater than the area increase of the group, and sometimes the same; her increase was significantly smaller than the group mean for only one region. She experienced the same alignment error on the ''bureau region'' as the

blindfolded-sighted participants, and when ranked among the other haptic participants, she fell among the top extenders, but was not an outlier. Although she was among the top extenders in the haptic condition, her boundary extension was always significantly smaller than the mean of the vision group (something, I might add, that KC found quite amusing). Clearly, boundary extension occurred in real space following visual inspection, haptic exploration (without vision), and exploration by a "haptic expert." Clearly, those who explored the scene using the visual modality made the largest errors. Why?

4.2. Cross-Modal Boundary Extension

There are several possible explanations of why visual exploration might lead to more expansive boundary extension. One possibility is that the difference does not reflect a difference in the representation, but instead reflects a bias during testing; people might simply have been more conservative in setting the boundaries at test manually when they could not see. In Intraub, Morelli, and Daniels (2009), we sought to determine if the difference between modalities observed in Intraub (2004) could be replicated using seven new stimulus regions; and then if so, to determine if the difference is due to the mode of exploration during perception, the mode of exploration at test, or both

We tested 80 participants in a 2 (input modality) × 2 (test modality) design. Participants perceived the regions using either vision or haptic exploration (without vision) and were then tested either using the same modality or the other modality. Boundary extension occurred for all seven regions, in all four conditions. Results revealed an effect of input modality, no effect of test modality and no interaction between the two. When the regions were explored visually, boundary extension was greater than when they were explored manually (without vision) irrespective of the test modality. This shows that boundary extension was unaffected by a cross-modal transfer at test. The decision about boundary location (perhaps as discussed earlier, a source monitoring decision) was influenced the modality used to originally perceive the region.

The direction of the effect (greater boundary extension in vision) can be explained in terms of the source monitoring account of boundary extension in the following way. The contact involved in manual exploration involves several high-acuity areas (five fingertips per hand) and a smaller peripheral region than vision. Perhaps as a result, memory for felt space differs more sharply from the amodally generated continuation of the region beyond the boundaries. If the difference is more discontinuous than in vision (single point of fixation and a very large periphery), this might result in a higher threshold for accepting amodally generated space as having been experienced though sensory input. Thus, this would result in less boundary

extension for the contact sense. This analysis certainly does not prove that source monitoring is involved, but is offered simply as a hypothesis that would be consistent with the multisource framework.

This suggests that boundary extension might be constrained, in part, by the nature of the input modality used to explore the world. During visual scanning, a small shift of the head and/or the eyes will bring a much larger new region into view than will a small shift in hand position during haptic exploration. In terms of the possible usefulness of boundary extension in anticipating upcoming layout, it would be more likely to help provide a sense of a continuous world if the extended region were large enough to facilitate integration with adjacent regions, but not so large as to be confusing or misleading given the characteristics of the input modality. Related to this, we found in another experiment in this series that when we allowed participants to use both vision and haptics simultaneously (bimodal input), boundary extension did not differ significantly from that obtained following haptic exploration alone. The haptic input apparently tempered the decision about where the edges of the region had been located.

4.3. Monocular Tunnel Vision and Boundary Extension

In two other experiments in this series, we sought to determine if the difference between the modalities reflected fundamental differences, or if the differences observed might actually be mediated by something relatively simple—the scope of each perceptual sample (i.e., the visual field is relatively large in comparison to the field of "view" associated with haptics). If the difference is simply related to the scope of each sample, then if we were to restrict the observer's field of view during visual exploration of the stimulus regions, we might be able to reduce the amount of boundary extension they experience. That is, restricted viewing might make memory following visual exploration more similar to that observed following haptic exploration.

To test this, we created vision blocking goggles, shown in Figure 9. There were two monocular tunnel vision conditions (large tunnel and small tunnel) in which vision was restricted, causing participants to have to move their heads to inspect each of three stimulus regions. A binocular viewing condition (as in the previous experiments) served as a baseline control. An illustration of presentation and test conditions for the small monocular tunnel condition is shown in Figure 10. Except for the fact that the two tunnel vision groups wore vision blocking goggles with peepholes during study and test, the procedure was the same as in the previous experiments. Observers with the large monocular tunnel view could position their heads so that they could see all or almost all of the region at one time (depending on window size), but still they had to move their heads around when they wanted to gain a high-acuity view of different parts of the region. Observers

Figure 9 Large monocular tunnel vision goggles (3 cm peephole) and small monocular tunnel goggles (0.6 cm peephole) from the tunnel vision experiments in Intraub, Morelli, and Daniels (2009).

Figure 10 An illustration of the presentation (left) and test procedure with the tunnel vision goggles (right) in the tunnel vision experiments in Intraub, Morelli, and Daniels (2009).

with the small monocular tunnel view could only see a fraction of the stimulus region at a time. They could see an entire object at once, only in the case of the smallest objects (miniature toy cars). These participants made frequent extreme shifts of head position to study the stimulus regions. Study time was 30 s in all conditions.

The results were somewhat surprising in that introduction of these extremely unnatural viewing had no effect on memory. All three scenes were remembered with extended boundaries in all three conditions. On average, observers increased the areas of the regions by about one-third their original size and no significant difference was obtained across the groups; the mean area increase for the binocular group, large monocular tunnel group, and small tunnel group was 32%, 31%, and 37%, respectively.

We conducted one more tunnel vision experiment for two reasons. First, we wanted to determine if a factor that affected boundary extension

with pictures would have a similar effect on boundary extension for these 3D views. We sought to determine if memory for more ''wide-angle views'' of the same objects and backgrounds (i.e., slightly larger window sizes) would yield less boundary extension. It is well established in the literature that wider-angle pictures that include more continuous background between the objects and the edges of the view elicit less boundary extension (e.g., Intraub et al., 1992). Second, to test the unlikely possibility that the three viewing conditions yielded the same amount of boundary extension because at test there was simply a favored location for placing the borders, we set the boundaries of the stimulus views in the new experiment to equal the mean positions provided by the previous participants. If this was a ''favored position'' then in setting the boundaries this time, no boundary extension would be expected.

Given the lack of a difference between the two tunnel conditions we tested only the two extremes (binocular viewing and small monocular tunnel viewing). Although the boundaries were set at the mean location chosen by the previous participants, boundary extension occurred in both conditions. Consistent with the results of picture studies, compared with the previous experiment, boundary extension was reduced for these ''wider views.'' The mean area increase was 7% in the binocular condition and 14% in the small monocular tunnel condition. Again, there was no significant difference between the two groups, and if anything, as in the previous experiment, the means favored greater boundary extension in the small monocular tunnel group rather than less (which would be expected if the limited spatial scope had made vision more like haptics).

4.4. Possible Clinical Implications

Simply reducing the size of each sample during exploration did not reduce boundary extension. The results suggest that it is the nature of the modality rather than the spatial scope of each sample that will affect memory. This outcome is interesting in light of a clinical observation made by ophthalmologists treating retinitis pigmentosa (a progressive eye disease in which the patient loses peripheral vision). There are some patients who when first reporting that they have an eye problem are surprisingly unaware that the problem involves major losses of peripheral vision—a phenomenon that has sometimes been attributed to denial (D. Lindsey, personal communication, December 9, 2005). Our results suggest a possible alternative. The lack of sensitivity to this dramatic deficit in the sensory input might in part reflect an intact multisource scene representation that masks the true nature of the patient's problem. Like our small tunnel participants, although of course much more gradually, these patients begin to search differently (increasing their head movements), but are still able to perceive and remember a

coherent, continuous scene; what is missing from peripheral vision is augmented by nonvisual sources in their multisource representation.

5. Summary and Conclusions

The multisource model presented here provides a possible framework for rethinking visual scene representation. According to this view, at its core, scene perception is an act of spatial cognition that builds upon a deeply maintained sense of surrounding space—as discussed here, an egocentric frame of reference (although other frameworks can be implemented as well; Epstein & Higgins, 2007). This reference frame is filled-in by vision, amodal perception (Kanizsa, 1979; Nakayama et al., 1995; Yin et al., 2000), and contextual information that is garnered through categorization of the global layout of a view (see Greene & Oliva, 2009) and through contextual associations that are triggered by the objects (Bar, 2004). The idea that associations evoke probable contexts is consistent in a broad sense with Bar's (2004) *multiplexer model*, although in the multisource model described here, these associations, along with the other sources of input, are organized within a surrounding spatial framework. The detailed quality and specificity of the representation is graded, just as the visual field itself is graded. The graded acuity present in visual and haptic input may serve to enhance incorporation of new sensory inputs into an active scene representation. Thus, our representation of a scene always goes beyond the sensory input. In the case of a single view, as discussed here, the representation becomes less constrained by specific visual information the farther from the boundary a location is. Thus, the boundary extension error is relatively small—only the most highly constrained region will be mistaken for having been seen (or touched in the case of haptics without vision).

Given the multisource nature of the representation, boundary extension can be explained by adopting, without change, a model designed to explain many long-term memory phenomena—the *source monitoring* model (Johnson et al., 1993; Lindsay, 2008). What is interesting is that in this context source monitoring can provide an explanation of a memory error that occurs following a very brief break in the sensory input ranging from a second or two (Bertamini et al., 2005; Intraub et al., 1996, 2006) down to intervals lasting less than 1/20th of a second (Dickinson & Intraub, 2008; Intraub & Dickinson, 2008). The notion here is that scene "extrapolation" does not take place after the stimulus is gone. Instead, the boundary extended region was already present in the scene representation while the visual input was available—in the form of amodal perception and contextual associations. It is only after a break in the sensory input, when the observer is

forced to monitor memory and to make a source attribution that highly constrained amodal information is attributed to having been seen.

Current visual processing models are by their nature *single-source* models and thus provide no predictions about scene representation based upon other modalities. In contrast, the multisource model has at its core a spatial frame of reference (in our discussion, an egocentric frame of reference) that is filled-in by multiple sources of input, with or without the inclusion of vision. This raises the expectation that there will be similarities in scene perception and memory across modalities. Such similarities have been observed with respect to boundary extension. Memory beyond the boundaries occurs following visual and haptic exploration (without vision) of the same 3D regions. Boundary extension also occurs when a background is imagined (no sensory modality; Intraub et al., 1998). A key point of the multisource framework is that whichever modality may be in the fore, scene representation will be a multisource representation that captures the continuity of layout in the world. This can be thought of as an example of situated perception (Barsalou, 1999) in which a view (a part of a scene) instantiates an encompassing scene representation.

Cognitive neuroscience research has provided interesting support for the notion of a multisource scene representation. This can be seen in research on the neural response to different types of questions about a picture that draw on different frames of reference (e.g., the neural activity associated with what is happening in the picture, e.g., "a party," the layout of the immediate view, or the integration of that view within a larger geographic framework; Epstein & Higgins, 2007). Evidence for the role of expected contexts caused by associations with objects has also been reported (Bar, 2004). This has resulted in an interesting controversy about whether or not PPA and RSC should be conceptualized as scene-selective regions as originally thought (e.g., Epstein, 2009; Epstein & Kanwisher, 1998) or as part of a network of more abstract conceptual associations (Bar, 2004). These are controversies that are yet to play out, and they raise interesting questions about scene perception and its relation to other aspects of cognition.

However, in the context of exploring objects in locations with the eyes or hands, I suggest that the underlying organizing structure that allows us to understand our world is not so much an *abstract* schema (Hochberg, 1986; Intraub, 1997) as a *concrete* sense of surrounding space in relation to the observer. The world is continuous, and surrounds us, we are embedded within an environment and navigate through it. Sensory input can never provide access to the details of our surroundings all at once—creating one of the classic puzzles that challenge theories of perception (e.g., Hochberg, 1986; O'Regan, 1992; Rensink, 2000). A multisource scene representation provides one possible explanation of how observers perceive a continuous world that they can sample only a part at a time; and why observers tend to remember seeing beyond the physical boundaries of a view just moments after that view is gone.

ACKNOWLEDGMENT

This work was supported in part by NIMH Grant MH54688.

REFERENCES

Allen, G. (2004). *Human spatial memory: Remembering where*. Mahwah, NJ: Erlbaum.

Atkins, J. E., Fiser, J., & Jacobs, R. A. (2001). Experience-dependent visual cue integration based on consistencies between visual and haptic percepts. *Vision Research*, *41*, 449–461.

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

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, *22*, 577–609.

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.

Biederman, I. (1981). On the semantics of a glance at a scene. In M. Kubovy & J. R. Pomerantz (Eds.), *Perceptual organization* (pp. 213–253). Hillsdale, NJ: Erlbaum.

Bryant, D. J., Tversky, B., & Franklin, N. (1992). Internal and external spatial frameworks for representing described scenes. *Journal of Memory and Language*, *31*, 74–98.

Chapman, P., Ropar, D., Mitchell, P., & Ackroyd, K. (2005). Understanding boundary extension: Normalization and extension errors in picture memory among normal adults and boys with and without Asperger's syndrome. *Visual Cognition*, *12*(7), 1265–1290.

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*, 543–555.

DiCola, C., & Intraub, H. (2004). *Reconstructing scenes: View reconstructing scenes: View-boundaries vs. boundaries vs. object object-boundaries boundaries*. Visual Science Society Meeting, Sarasota, FL.

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

Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature*, *392*(9), 598–601.

Epstein, R. Al., & Ward, E. J. (2009). How reliable are visual context effects in the parahippocampal place area? *Cerebral Cortex*, Advance Access published on June 16, 2009; doi:10.1093/cercor/bhp099.

Fantoni, C., Hilger, J. D., Gerbino, W., & Kellman, P. J. (2008). Surface interpolation and 3D relatability. *Journal of Vision*, *8*(7)doi:10.1167/8.7.29 29, 1–19, http://journalofvision.org/8/7/29/.

Franklin, N., & Tversky, B. (1990). Searching imagined environments. *Journal of Experimental Psychology: General*, *119*, 63–76.

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.

Gottesman, C. V., & Intraub, H. (2003). Constraints on spatial extrapolation in the mental representation of scenes: View-boundaries vs. object-boundaries. *Visual Cognition*, *10*(7), 875–893.

Greene, M. R., & Oliva, A. (2009). Recognition of natural scenes from global properties: Seeing the forest without representing the trees. *Cognitive Psychology*, *58*, 137–176.

Henderson, J. M., & Hollingworth, A. (1999). High-level scene perception. *Annual Review of Psychology*, *50*, 243–271.

Henderson, J. M., & Hollingworth, A. (2003). Eye movements and visual memory: Detecting changes to saccade targets in scenes. *Perception & Psychophysics*, *65*, 58–71.

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. 22.1–22.64). New York: Wiley.

Intraub, H. (1984). Conceptual masking: The effects of subsequent visual events on memory for pictures. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *l0*, 115–125.

Intraub, H. (1997). The representation of Visual Scenes. *Trends in the Cognitive Sciences*, *1*, 217–221.

Intraub, H. (2002). Anticipatory spatial representation of natural scenes: momentum without movement?. *Visual Cognition*, *9*, 93–119.

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

Intraub, H. (2007). Scene perception: The world through a window. In M. A. Peterson, B. Gillam & H. A. Sedgwick (Eds.), *The mind's eye: Julian Hochberg on the perception of pictures, films, and the world* (pp. 454–466). New York: Oxford University Press.

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*, 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*, 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., Gottesman, C. V., Willey, E. V., & Zuk, I. J. (1996). Boundary extension for briefly glimpsed photographs: Do common perceptual processes result in unexpected memory distortions? *Journal of Memory and Language*, *35*, 118–134.

Intraub, H., Hoffman, J. E., Wetherhold, C. J., & Stoehs, S.-A. (2006). More than meets the eye: The effect of planned fixations on scene representation. *Perception & Psychophysics*, *68*(5), 759–769.

Intraub, H., Morelli, F., & Daniels, K. K. (2009). Exploring the world by eye and by hand. Manuscript in preparation.

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*, 420–456.

Irwin, D. E. (1993). Perceiving an integrated visual world. In D. E. Meyer & S. Kornblum (Eds.), *Attention and performance 14: Synergies in experimental psychology, artificial intelligence, and cognitive neuroscience* (pp. 121–142). Cambridge, MA: MIT Press.

James, W. (1890). *The principles of psychology*. Vol. II New York: Holt and Company.

Johnson, M. K. (2006). Memory and reality. *American Psychologist*, *61*, 760–771.

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

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

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

Koriat, A., Goldsmith, M., & Pansky, A. (2000). Toward a psychology of memory accuracy. *Annual Review of Psychology*, *51*, 481–537.

Legault, E., & Standing, L. (1992). Memory for size of drawings and of photographs. *Perceptual and Motor Skills*, *75*, 121.

Lindsay, D. S. (2008). Source monitoring. In J. Byrne (Ed.), *Cognitive psychology of memory. Vol. 2 of learning and memory: A comprehensive reference, 4 vols* (pp. 325–348). Oxford: Elsevier.

Loftus, G. R., & Ginn, M. (1984). Perceptual and conceptual processing of pictures. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *10*, 435–441.

Loftus, G. R., Johnson, C. A., & Shimamura, A. P. (1985). How much is an icon worth? *Journal of Experimental Psychology: Human Perception and Performance*, *11*, 1–13.

Michod, K., & Intraub, H. (2009). Boundary extension. *Scholarpedia*, *4*(2), 3324.

Nakayama, K., He, Z. J., & Shimojo, S. (1995). Visual surface representation: A critical link between lower-level and higher level vision. In S. M. Kosslyn & D. N. Osherson (Eds.), *Invitation to Cognitive Science* (pp. 1–70). Cambridge, MA: MIT Press.

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

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

O'Regan, J. K., & Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences*, *24*, 939–1011.

Park, S. J., 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.

Phillips, W. A. (1974). On the distinction between sensory storage and short-term visual memory. *Perception & Psychophysics*, *16*(2), 283–290.

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

Potter, M. C. (1999). Understanding sentences and scenes: The role of conceptual short-term memory. In V. Coltheart (Ed.), *Fleeting memories: Cognition of brief visual stimuli* (pp. 13–46). Cambridge, MA: MIT Press.

Previc, F. H. (1998). The neuropsychology of 3-D space. *Psychological Bulletin*, *124*, 123–164.

Quinn, P. C., & Intraub, H. (2007). Perceiving ''outside the box'' occurs early in development: Evidence for boundary extension in 3- to 7-month-old infants. *Child Development*, *78*, 324–334.

Rensink, R. A. (2000). The dynamic representation of scenes. *Visual Cognition*, *7*(1–3), 17–42.

Rensink, R. A., O'Regan, J. K., & Clark, J. J. (1997). To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science*, *8*, 368–373.

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

Simons, D. J., Franconeri, S. L., & Reimer, R. L. (2000). Change blindness in the absence of visual disruption. *Perception*, *29*, 1143–1154.

Simons, D. J., & Rensink, R. A. (2005). Change blindness: Past, present, and future. *Trends in Cognitive Sciences*, *9*(1), 16–20.

Standing, L. (1973). Learning 10, 000 pictures. *The Quarterly Journal of Experimental Psychology*, *25*, 207–222.

Standing, L., Conezio, J., & Haber, R. N. (1970). Perception and memory for pictures: Single-trial learning of 2500 visual stimuli. *Psychonomic Science*, *19*, 73–74.

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, *381*(6), 520–522.

Turk-Browne, N. B., Scholl, B. J., & Chun, M. M. (2008). Habituation in infant cognition and functional neuroimaging. *Frontiers in Human Neuroscience*, *2*, 16.

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

VanRullen, R., & Thorpe, S. J. (2001). Is it a bird? Is it a plane? Ultra-rapid visual categorisation of natural and artifactual objects. *Perception*, *30*, 655–668.

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

Zelinsky, G. J., & Loschky, L. C. (2005). Eye movements serialize memory for objects in scenes. *Perception & Psychophysics*, *67*, 676–690.

Zelinsky, G. J., & Schmidt, J. (2009). An effect of referential scene constraint on search implies scene segmentation. *Visual Cognition*, *17*, 1004–1028.