Cognition 138 (2015) 132–147

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

Visual, haptic and bimodal scene perception: Evidence for a unitary representation

Helene Intraub *, Frank Morelli [1], Kristin M. Gagnier [2]

University of Delaware, USA

ARTICLE INFO

Article history:
Received 3 July 2014
Revised 21 January 2015
Accepted 25 January 2015
Available online 25 February 2015

Keywords:
Scene representation
Spatial memory
Multisensory (visual and haptic)
Cross-modal representation
Boundary extension

ABSTRACT

Participants studied seven meaningful scene-regions bordered by removable boundaries (30 s each). In Experiment 1 ($N = 80$) participants used visual or haptic exploration and then minutes later, reconstructed boundary position using the same or the alternate modality. Participants in all groups shifted boundary placement outward (boundary extension), but visual study yielded the greater error. Critically, this modality-specific difference in boundary extension transferred without cost in the cross-modal conditions, suggesting a functionally unitary scene representation. In Experiment 2 ($N = 20$), bimodal study led to boundary extension that did not differ from haptic exploration alone, suggesting that bimodal spatial memory was constrained by the more "conservative" haptic modality. In Experiment 3 ($N = 20$), as in picture studies, boundary memory was tested 30 s after viewing each scene-region and as with pictures, boundary extension still occurred. Results suggest that scene representation is organized around an amodal spatial core that organizes bottom-up information from multiple modalities in combination with top-down expectations about the surrounding world.

1. Introduction

Multiple sensory modalities provide the perceiver with rich information about the surrounding world. In spite of this, similar to other areas of perception, research on *scene perception* has typically been studied through a modality-specific lens (usually vision; Intraub, 2012; O'Regan, 1992). Yet, even when perception is limited to the visual modality alone participants frequently remember seeing the continuation of the scene just beyond the boundaries of the view, in the absence of any corresponding sensory input (boundary extension; Intraub & Richardson, 1989). This can occur very rapidly, across intervals as brief as a saccadic eye movement (Dickinson & Intraub, 2008; Intraub & Dickinson, 2008). Boundary extension may be an adaptive error that facilitates integration of successive views of the world (Hubbard, Hutchison, & Courtney, 2010; Intraub, 2010, 2012). Indeed, research has shown that boundary extension can prime visual perception of upcoming layout, when that layout is subsequently presented (e.g., Gottesman, 2011).

What leads to this spatial error? Intraub (2010, 2012) and Intraub and Dickinson (2008) suggested that rather than a *visual representation*, representation of visual scenes is actually a *multisource* representation in that it incorporates information from both the sensory source (vision) as well as top-down sources of information that place the studied view within a likely surrounding spatial context. Potential top-down sources include amodal continuation

* Corresponding author at: Department of Psychological and Brain Sciences, University of Delaware, Newark, DE 19716, USA.

E-mail address: intraub@udel.edu (H. Intraub).

[1] Address: Dismounted Warrior Branch, Human Research and Engineering Directorate, U.S. Army Research Laboratory, Aberdeen Proving Ground, MD 21005, USA.

[2] Address: Spatial Intelligence and Learning Center, Department of Psychology, Temple University, Philadelphia, PA 19122, USA.

http://dx.doi.org/10.1016/j.cognition.2015.01.010
0010-0277/

of the surface beyond the boundaries (Fantoni, Hilger, Gerbino, & Kellman, 2008), general scene knowledge based upon scene classification (Greene & Oliva, 2009), and object-to-context associations (Bar, 2004). The purpose of our research was to determine if boundary extension following visual or haptic perception of the same scene-region is supported by a single *multimodal* scene representation or by two functionally independent modality-specific scene representations.

Boundary extension is a spatial error in which a swath of anticipated space just beyond the boundaries of the view is remembered as having been perceived. Neuroimaging and neuropsychological research have shown that boundary extension is associated with neural activation of brain regions thought to play important roles in spatial cognition: the hippocampus, parahippocampal cortex, and retrosplenial complex (Chadwick, Mullally, & Maguire, 2013; Mullally, Intraub, & Maguire, 2012; Park, Intraub, Yi, Widders, & Chun, 2007). The hippocampus has long been associated with spatial representation and navigation (Burgess, 2002; Maguire & Mullally, 2013; O'Keefe & Nadel, 1978). The parahippocampal cortex and retrosplenial complex have been associated with perception of spatial layout, and with the integration of local spaces within larger spatial contexts, respectively (Epstein, 2008). Recent research has shown that the parahippocampal cortex responds similarly to visual and haptic perception of layout (Wolbers, Klatzky, Loomis, Wutte, & Giudice, 2011; also see Epstein, 2011), underscoring the spatial rather than modality-centric role of this brain area.

It has been suggested that scene representation is fundamentally an act of spatial cognition (Dickinson & Intraub, 2008; Gagnier, Dickinson & Intraub, 2013; Gagnier & Intraub, 2012; Intraub, 2010, 2012; Intraub & Dickinson, 2008). In their *multisource model* Intraub and Dickinson (2008; Intraub, 2010, 2012) proposed that an amodal spatial structure organizes multiple sources of knowledge (bottom-up and top-down) into a coherent scene representation (see Maguire & Mullally, 2013, for a similar view from the perspective of hippocampal function). The idea is that the observer brings to any view of a scene a sense of surrounding space (the space "in front of", "to the left and right", "above", "below" and "behind" the observer (see Bryant, Tversky, & Franklin, 1992; Franklin & Tversky, 1990; Tversky, 2009). This provides the scaffolding that supports not only the bottom-up visual information but the anticipated continuation of the scene beyond the boundaries of the view. This underlying spatial structure is similar to the "spatial image" proposed by Loomis, Klatzky, and Giudice (2013), in that it is a surrounding spatial representation (not limited to the frontal plane) and that it is amodal. The only difference is that unlike the "spatial image", the spatial structure in the multisource model is conceptualized as a "standing framework", rather than one that develops in response to a stimulus in working memory.

Although most research on boundary extension has focused on picture memory, there is evidence that the same anticipatory spatial error occurs following visual perception of real scenes in near space (Hubbard et al., 2010; Intraub, 2002, 2010), and following haptic perception of the same scene regions (Intraub, 2004; Mullally et al., 2012). The multisource model provided the same explanation for visual and haptic boundary extension, but included no commitment as to whether they draw on a single scene representation or on distinct modality-specific representations. The evidence for boundary extension in 3D space was based on experiments in which meaningfully related objects were arranged on natural backgrounds (e.g., "kitchen scene"), bounded by a "window frame" to limit visual or haptic exploration.

In haptic studies, blindfolded participants explored the bounded regions right up to edges of the display, and minutes later, after the boundaries were removed, participants reconstructed boundary placement. They set the boundaries outward, including a greater expanse of space that had originally been included in the stimulus. This occurred in spite of the fact that there was always an object 2–3 cm from the boundary, forcing participants to squeeze their hands into a tightly constrained space. As in the case of vision (Gagnier et al., 2013) a seemingly clear marker of boundary placement did not prevent boundary extension. A comparison of boundary extension following visual or haptic exploration of the same regions showed that vision yielded the more expansive error (Intraub, 2004). This was the case whether visual boundary extension was compared to haptic boundary extension in sighted participants who were blindfolded for the experiment, or in a woman who had been deaf and blind since early life (a "haptic expert").

Why might vision have yielded a greater anticipatory spatial error? Intraub (2004) speculated that such a difference, if reliable, might be related to the different characteristics and spatial scope of the two modalities. Vision is a distal modality with a small high acuity foveal region (about 1° of visual angle) and a large low-acuity periphery. Together these encompass a relatively large spatial area. In contrast, the haptic modality encompasses multiple high acuity regions (the fingertips) and a relatively small periphery. In the case of vision, a greater amount of the visually imagined continuation of the view might be confusable with visual memory for the stimulus than in the case of haptic exploration. This explanation conforms to the notion of boundary extension as a source monitoring error (Intraub, 2010, 2012; Intraub, Daniels, Horowitz, & Wolfe, 2008; Intraub & Dickinson, 2008; Seamon, Schlegel, Hiester, Landau, & Blumenthal, 2002). According to the source monitoring framework (Johnson, Hashtroudi, & Lindsay, 1993) as the similarity between representations drawn from two different sources (e.g., perception and imagination) increases, so too does the likelihood of source misattributions (as, for example, when a dream is unusually high in detail, and is erroneously misattributed to perception). Other research on boundary extension has shown that factors that would be expected to affect the similarity between memory for the perceived region and memory for the imagined continuation of the view do indeed influence the size of the boundary error (Gagnier & Intraub, 2012; Intraub et al., 2008). An alternative explanation of the difference between the visual and haptic conditions, however, is that it is caused by different biases at test. For example, blindfolded participants may feel more constrained about how far they are comfortable reaching out their hands to designate boundary placement.

In sum, both visual and haptic exploration can result in boundary extension. However, the studies demonstrating this cannot provide insight into whether the scene representation supporting this error is a single multisource scene representation that includes multimodal input, or two separate, modality-specific multisource representations, with *multisource* referring to the combination of bottom-up and top-down sources of information. Before describing the rationale for our research, we will discuss other spatial tasks in which visual and haptic study have been compared, because they have direct bearing on the current research.

A critical aspect of spatial cognition is that an arrangement of objects in the world can be represented within a variety of reference frames (see Allen, 2004). A well-established observation is that after viewing a display of objects, participants tend to organize memory within an egocentric frame of reference (the objects with respect to the viewer), rather than an allocentric framework (the objects with respect to one another; Diwadkar & McNamara, 1997; Shelton & McNamara, 1997; Simons & Wang, 1998; Wang & Simons, 1999). Critical observations supporting this are that costs are incurred when participants either view or imagine the display from an alternate viewpoint (e.g., a viewpoint that is shifted 60° from the original position). In these cases, when the opportunity for spatial updating is eliminated (e.g., as would occur if participants simply walked to the new location) the change in viewpoint reduced participants' ability to remember the object arrangement correctly. Haptic exploration, unlike vision (with its much larger periphery), does not allow the observer to perceive all the objects at once. Instead the observer must serially explore each object's relation to other objects individually, raising the possibility that an allocentric representation might be engaged (Newell, Woods, Mernagh, & Bülthoff, 2005). However, analogous research using the haptic modality yielded similar results (Newell et al., 2005; Yamamoto & Shelton, 2005). Spatial memory under these conditions was organized within an egocentric frame of reference when the scene was perceived using haptic input.

In subsequent frame-of-reference research in the visual modality, Mou and McNamara (2002) demonstrated that if the display of objects is arranged so that it has an intrinsic structure (e.g., symmetry), rather than an egocentric reference frame, participants adopt an allocentric frame of reference that is centered on the intrinsic structure of the display. Yamamoto and Philbeck (2013) pointed out that the smaller scope of haptic exploration (compared with vision) might prevent participants from organizing memory around this intrinsic structure (which in haptics would require an object by object search). However, they reported that although the layout of the objects could not be perceived all at once, memory following haptic exploration mirrored that observed in vision. Participants organized their memory around the intrinsic structure of the displays. There is also evidence that environmental information provided by haptics can impact the choice of reference frame in a visual task, supporting the idea of a common reference frame across modalities (Kelly, Avraamides, & Giudice, 2011). Thus, results suggested that vision and haptics both share common biases in terms of reference frame, and both may support computation of the anticipated continuation of the explored region, yielding boundary extension in memory. In addition, other spatial tasks involving visual or haptic exploration of maps or scenes have demonstrated common representational biases (e.g., Giudice, Betty, & Loomis, 2011; Pasqualotto, Finucane, & Newell, 2005). These studies suggest that vision and haptics share common biases in terms of reference frame and Intraub's (2004) research adds to these commonalities, in demonstrating that both modalities yield the same overinclusive anticipatory spatial bias in memory – boundary extension.

Bearing these similarities in spatial biases in mind, we will now return to the question of whether visual information and haptic information are stored in a functionally unitary mental structure or in two functionally distinct, modality-specific representations. This has been a core issue in the field of multisensory perception. Different tasks and different types of stimuli have led to a variety of conclusions about how the sensory systems interact. Some evidence supports a single (multisensory) representation, whereas other evidence such as visual capture (Hay, Pick, & Ikeda, 1965) and auditory capture (Morein-Zamir, Soto-Faraco, & Kingstone, 2003) suggest modality-specific representations that in some cases yield conflicting information about stimuli. Models have been proposed to describe the different combination strategies and mechanisms used to integrate multisensory information into a coherent representation (Ernst & Bülthoff, 2004).

Multisensory studies have focused both on the temporal integration of independent objects (e.g., a visual stimulus and a sound) and on spatial integration (multiple modalities exploring the same object or display). The current research focuses on the spatial representation of meaningful scenes that are perceived visually, haptically (without vision), or bi-modally. To achieve this, stimuli were meaningful, multi-object displays (e.g., a place setting; tools in a workman's area) in near space directly in front of the participant (*peripersonal* space, Previc, 1998) that can be readily explored either using the visual or haptic modality.

Newell et al. (2005) addressed the question of a unitary representation vs. modality-specific representations in memory for object arrays in peripersonal space by contrasting recognition memory (detecting that the position of two of seven small wooden objects was swapped) within vision and within haptics vs. across modalities. They provided examples of cases in which different spatial biases were observed in vision and haptics (e.g., in the case of the horizontal-vertical illusion; Avery & Day, 1969; Day & Avery, 1970) and argued that given these differences, it may be that modality-specific representations would be formed in their experiment. They reasoned that this would result in a cost being incurred when recognition memory is tested across modalities because the information in one modality-specific representation would need to be "translated" into the other. On the other hand, if memory for spatially arrayed objects is maintained in a functionally unitary spatial representation, then no cost would be expected. Participants should be able to note differences in the positions of the objects irrespective of whether the modalities at study and test were the same or different. They tested participants' ability to recognize changes in object position within modality or across modalities, and did so whether the table was in the same position or was shifted 60° while the participant's view was blocked.

Newell et al. (2005) found no costs associated with cross-modal transfer as a function of viewpoint (*same view* vs. *shifted view*), suggesting that the same egocentric representation was supporting both visual and haptic representation (also see Kelly & Avraamides, 2011), but they did find a cost in recognition memory for the objects' positions as a function of modality. Participants made fewer errors in the within-modality conditions than the cross-modal conditions. Newell et al. argued that recoding spatial position from one representation to the other had incurred the cost. They argued that it was the spatial placement of the objects with respect to one another, rather than the specific details of how the objects themselves were remembered that was driving the difference, although this distinction was not specifically tested. In conclusion, Newell at al. suggested that whereas frame of reference is a unitary representation, specific spatial characteristics within that reference frame may be stored in different modality-specific representations.

We report three experiments in which spatial memory for the expanse of scene-regions composed of meaningfully related objects was examined following visual inspection, haptic exploration or both simultaneously (bi-modal exploration). In Experiment 1, similar to Newell et al. (2005), tests involved either the same modality or the alternative modality (testing cross-modal transfer). In Experiment 2, simultaneous bimodal exploration was used. In Experiment 3, we explored the possibility that boundary extension in peripersonal space requires navigation prior to testing. What is different about our research on spatial memory is that the focus is on *false* memory beyond the scope of the sensory input. Key questions explored across these three experiments were: (a) Is scene representation a unitary representation that incorporates information from multiple sensory sources (vision and haptics) with related top-down information or is information stored in separate modality-specific representations (one for vision and one for haptics) with each including associated top-down information, (b) If a unitary representation is supported, is there a "blending" of inputs into a code that is devoid of sensory specific characteristics (an amodal code) or does the mental representation retain qualities specifically tied to the individual modalities? And, (c) Can we observe boundary extension in memory for 3D scene regions under conditions that are similar to those that have been observed in memory for 2D scenes (photographs)?

2. Experiment 1

Participants were assigned to one of four independent groups in which study and test were conducted within-modality (*vision–vision,* or *haptic–haptic*) or across-modalities (*vision–haptic, or haptic–vision*). If memory is stored in a functionally unitary representation, then any modality-specific differences in boundary extension (e.g., greater boundary extension for visual than haptic exploration; Intraub, 2004) should transfer (without cost) in the cross-modal test conditions. We should see an effect of input modality rather than of test modality. If memory is stored in modality-specific representations, then the amount of boundary extension would be expected to differ depending on how it is tested, reflecting the costs of transferring between codes. Finally, if the difference in boundary extension between the vision and haptic conditions is due, not to encoding, but to biases associated with each modality at test, then we should observe an effect of test modality but no effect of input modality.

Because participants were not allowed to move the objects, it is possible that changes in boundary placement could be due to participants misremembering the placement of the objects rather than the size of the region. For example, if they remembered the objects as being closer together than what they saw at test, to compensate, they would move the boundaries outward. To test this possibility, following Intraub's (2004) procedure, participants were asked to describe any changes in object position they noted in the test stimuli. Finally, we should point out that all objects differed from those in Intraub (2004) so we could determine if the basic observations reported in that study would be replicated with a new stimulus set. Six scene-regions were similar in that they were arrayed horizontally (on the floor and on tabletops), and a seventh display was added that was arrayed vertically on a surface perpendicular to the floor (on the front of a tall cabinet), to test the generality of boundary extension to different surface orientations.

2.1. Method

2.1.1. Participants

Participants were 80 (38 female) University of Delaware undergraduates, fulfilling a requirement for an introductory psychology course ($N = 20$ in each condition). It was decided *a priori*, that because the seven scene-regions differed in size and aspect ratio, we would replace any participants who had missing data; we replaced one participant in the *vision–vision* condition and four in the *haptic–vision* condition because they created one or more highly distorted scene regions when setting the boundaries (e.g., a triangular scene region) so that meaningful boundary measurements could not be taken on those trials. Also, to avoid an effect of outliers, we set an exclusion rule for participants who created areas 3 SDs or greater from the mean; no such exclusions were necessary in Experiment 1.

2.1.2. Stimuli and apparatus

Stimuli consisted of seven scene-regions, each bounded by a frame. The regions and their dimensions are shown in Fig. 1. Two stimuli ("*place setting*" and "*baby scene*") were in a room that was $23' \times 9'$, and the remaining five stimuli were in another room that was $19' \times 18'$, connected by a small hallway with a chair that served as a "waiting area." Scene-regions were positioned so that vision participants

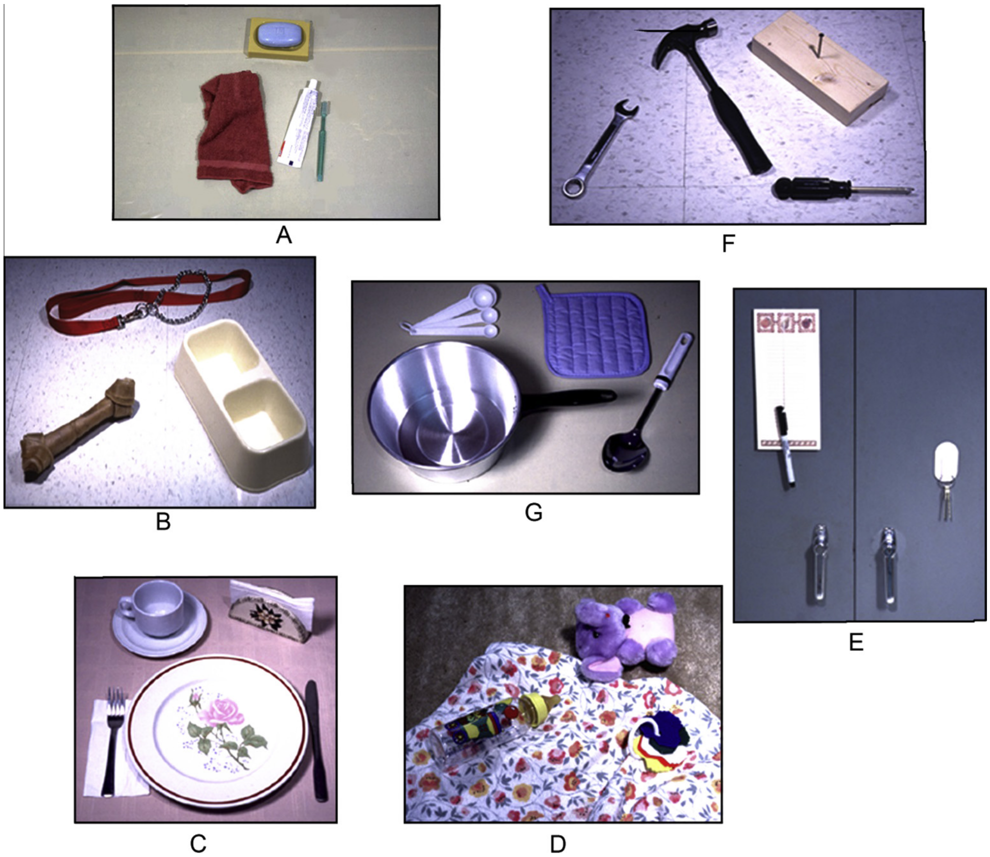

Fig. 1. The seven scene-regions in order of presentation: (a) Sink: 20″ × 17″ (51 cm × 43 cm), (b) dog: 24″ × 24″ (61 cm × 61 cm), (c) setting: 15″ × 19″ (38 cm × 48 cm), (d) baby: 18″ × 18″ (46 cm × 46 cm), (e) cabinet: 17″ × 23″ (43 cm × 58 cm), (f) tools: 19″ × 14″ (48 cm × 36 cm), (g) kitchen 22″ × 16″ (56 cm × 41 cm).

could see only one scene region at a time. Adhesive putty fixed the objects in place to avoid accidental displacement during haptic exploration; for the one vertical scene (in which the background was the large metal door of a tall cabinet; see Fig. 1) magnetic strips glued to the backs of the objects and the boundaries kept them in place.

Two different types of frames were used for the visual and haptic study conditions because of the different nature of the modalities. For visual study, it was important that the frame lay flat on the floor so that the edges would not occlude part of the view. Also it was critical to cover the surrounding space just outside the boundaries so that participant's peripheral vision would not include visual information about the studied surface beyond the boundaries. In the visual study conditions, the bounding frame was a rectangle constructed of flat wooden strips to which black cloth was attached (as shown in Fig. 2, left panel). There was therefore a stark change in the visual information inside and outside the boundaries, and information about the visual appearance of the surface relative to the boundary was only available within the bounded area.

For haptics, the flat frame used for vision does not create a sufficient boundary to prevent participants from accidently moving their hands outside the stimulus region. For this reason, as in Intraub (2004), we created taller boundaries for the haptic condition, as this does not "occlude the view" but does help the participant stay on task. Because participants in this condition would be blindfolded, we did not include black cloth outside the perimeter. In the haptic study conditions, the frame was 3″ (7.62 cm) tall, and was supported by an external frame that the participant never touched (as shown in Fig. 2, right panel). The exception to these two types of apparatus was the frame used in the single vertical scene region. This frame was made of Styrofoam strips (with cloth attached in the visual condition), held in place with magnetic strips. Thus, the visual and haptic conditions were made to be a similar as possible given the differences in the nature of each modality – the scene region was clearly defined and set off from the surrounding space. At test, in the visual test conditions, individual flat wooden strips (or Styrofoam in the case of the vertical scene) with cloth were used to reconstruct boundary placement. In the haptic test conditions, heavy markers (bricks wrapped with duct tape) were used to mark the remembered boundary location. The smooth surface of the sticks (visual test) and the taped bricks (haptic test) allowed participants to slide the markers to make even very slight adjustments in their position.

Fig. 2. Visual exploration (left) and haptic exploration (right) of the "dog" scene. (All boundaries were removed prior to test.)

Vision blocking goggles were plastic safety goggles (commonly used in chemistry labs) that were painted black and covered with duct tape (because the paint was easily chipped). In addition, sterile gauze was placed in each nose piece to prevent peeking through that area. Goggles were sterilized and gauze discarded after each use. A stopwatch was used to time the 30 s exploration time in both conditions.

2.1.3. Design and procedure

Participants were run individually and were randomly assigned to one of the four conditions (*vision–vision, vision–haptic, haptic–haptic, haptic–vision*). Because of the cumbersome nature of the equipment, a different condition was setup each day, and condition was counterbalanced across days, with approximately five participants per day (depending on the show rate). This continued until all four conditions were complete. All participants donned vision-blocking goggles in a waiting area prior to entering the test rooms. In this way, vision participants and haptic participants were limited to the same "view" of each scene region during the 30 s study period. All participants experienced the scene-regions from the same fixed location, kneeling on a small carpet or standing by the edge of a tabletop display, or standing directly in front of the single vertical display. Again, the surrounding space was blocked for vision participants with black cloth, so that the two conditions would be as similar as possible.

2.1.3.1. Study instructions. All participants were instructed to look at (or feel) the entire scene-region, right up to the boundaries, trying to remember it in as much detail as possible. They were told to remember all the objects and their layout within the scene-region. Visual-study participants were instructed to restrict their gaze to the stimulus region and haptic-study participants were instructed to restrict hand movements to the stimulus region. The importance of this was underscored and explained (both groups were told we were comparing vision and touch and needed to be sure that all participants studied the same regions). The experimenter and assistants observed participant behavior. Participants followed these instructions.

Haptic-study participants were instructed to use both hands to ensure good object identification (see Klatzky, Lederman, & Reed, 1987), and were asked to touch the objects lightly to avoid dislodging them. The experimenter watched the placement of each participant's hands and reminded them to use both hands to when touching the objects and to explore the space up to the edges of the boundaries. Participants in both conditions were instructed to continue to study the scene region until the experimenter told them to stop at the end of the timed study period.

All participants were required to verbally describe the objects in the displays during study (this was done to ensure that participants in the haptic condition could correctly identify the objects; and that the same verbal activity would take place in the visual condition). After the 30-s study period was over, participants in the visual-study condition replaced their goggles. Participants were then instructed to stand up and to provide a short title to convey the general gist of the scene-region. With vision-blocking goggles in place, each participant was then led to the next scene region (in the same order) and the procedure was repeated until all seven regions were explored. At this point, participants were escorted to the waiting area where they were allowed to remove their goggles and were read the instructions for the test phase (5 min). During this time, two assistants quickly removed the frames from around the scene-regions in the test rooms.

2.1.3.2. Test instructions. Participants were informed that they would re-explore each display from the same position (either using the same modality or the alternate modality, depending upon their group assignment). They were asked to use a fingertip to signify the remembered location of each boundary, and to keep their finger in place until the assistant placed a marker at that location. Once the boundaries were in place, participants were allowed to make any fine adjustments and then rated their confidence ("sure", "pretty sure" or "not sure"). Again, all participants were blindfolded throughout the test phase, except when the vision participant slipped their goggles up to specify boundary placement at each scene region. The experimenter always looked down when placing the boundaries at the designated location to avoid eye contact with vision participants, and if the participant indicated a change, the experimenter always asked which way (inward or outward) to avoid experimenter bias.

After reconstructing boundary placement in all 7 regions, participants were asked to describe any changes in object placement they noted during the test (in actuality, no objects had been moved). Once the session ended and the participant left the room, boundary placement was measured as described in the next section.

2.1.3.3. Boundary measurement. When we setup the scenes, we also set markers to allow us to measure original boundary placement and the participant's remembered boundary placement. For tabletop scenes, the edges of the table served as the marker. For the vertical scene, the edges of the cabinet door served as the markers. For the floor scenes, small pencil marks were made parallel to each boundary at a point beyond which boundary extension was unlikely. Because the floor tiles were mottled, these small marks were well camouflaged. None were visible from the participant's location. The original placement of each boundary and the participant's remembered placement of the boundary was measured with respect to these reference points after the session was completed and the participant left the lab. The experimenter and the assistant reviewed each measurement to minimize error. To double check measurements made to these reference points, the assistants also made direct measurements of the length and width of each reconstructed "frame". On average, measurements derived using the reference points and direct measures of length and width fell within 2 mm of one another.

Participants' boundary placement (in terms of stimulus area, length, width, and distance of each boundary from the center of the scene-region) is reported as a percentage of the original. Thus, for example, if the reconstructed *area* was the same as the stimulus area, the percent width remembered would be 100%, if the boundaries were shifted outward the percentage would surpass 100% and if shifted inward, it would be less than 100%.

2.2. Results

All participants readily identified every object and agreed about the "gist" of the scene regions, providing similar titles, such as, "kitchen scene". Mean confidence rating was "pretty sure (2)" and did not differ significantly across conditions; mean ratings were 2.0 (SD = 0.6), 2.1 (SD = 0.5), 2.0 (SD = 0.6) and 2.0 (SD = 0.5) in the *vision–vision, haptic–haptic, vision–haptic* and *haptic–vision* conditions, respectively.

2.2.1. Remembered area

Fig. 3 shows the mean percent area remembered in each condition along with the .95 confidence intervals: as shown in the graph, significant boundary extension occurred in every condition. (In terms of individual mean percent area remembered, the tendency to move the boundaries outward was observed for 78 of the 80 participants We conducted a 2 × 2 × 7 mixed measures ANOVA, *input modality* (vision vs. haptic) × *test modality* (vision vs. haptic) × *scene-region* (the seven scenes).This allowed us examine the effects of input modality and test modality on boundary extension while at the same time

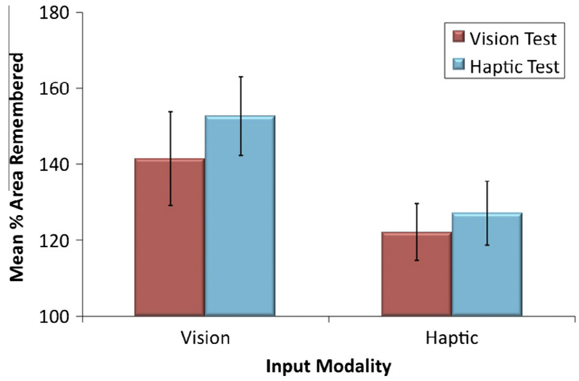

Fig. 3. Mean area remembered (%) in the visual and haptic input modality conditions as a function of the test modality (Experiment 1). Error bars indicate the 95% confidence interval for each mean.

determining if these effects interacted with individual scenes. Visual exploration led to greater boundary extension than did haptic exploration, $F(1,76) = 22.79$, $p < .001$, $d = 1.08$ and this difference did not interact with test modality, $F < 1$. In fact, there was no significant effect of test modality on performance, $F(1,76) = 2.94$, $p = .09$, $d = .34$, and the mean clearly showed no tendency toward greater boundary extension for visual tests than haptic tests.

The upper panel of Fig. 4 shows the mean percent area set by participants in the visual input groups (*vision–vision* and *vision–haptic*), and the lower panel of the figure shows

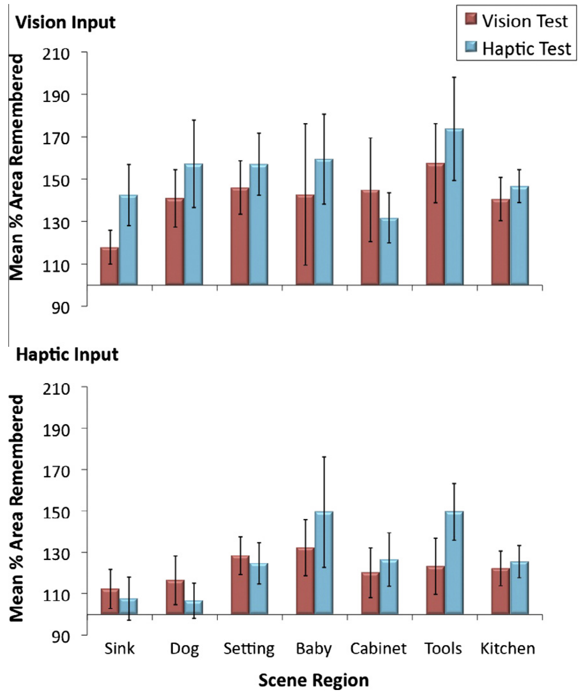

Fig. 4. Mean area remembered (%) for each scene-region when the input modality was visual (upper panel) and when the input modality of haptic (bottom panel) as a function of test modality (Experiment 1). Error bars indicate the 95% confidence interval for each mean.

the mean percent area set by participants in the haptic input groups (*haptic–haptic* and *haptic–vision*). As may be seen in the figure, in all cases the means were greater than 100% and the .95 confidence intervals show that in only 2 of the 28 comparisons did the increase in area fail to reach significance. The 3-way ANOVA showed that the amount of boundary extension differed across scene regions, $F(6,456) = 10.71$, $p < .001$; a common observation in research on boundary extension in memory for photographs. There is not as yet a clear explanation for this, and it is likely due to multiple factors. For this reason we had no predictions regarding higher level interactions and present them here from completeness. Scene × input modality approached significance, $F(6,456) = 2.04$, $p = .06$; scene × test modality, $F(6,456) = 1.84$, $p = .09$; and scene × input modality × test modality, $F(6,456) = 2.0$, $p = .06$.

2.2.2. Remembered length and width

Area changes can mask differences that might occur in length vs. width, so we conducted the same analyses in each dimension independently. The mean percent change in length and width in all four conditions is shown in Fig. 5, and results directly paralleled what was observed for area. A 2 (input modality) × 2 (test modality) ANOVA on the mean percent width revealed an effect of input modality, $F(1,76) = 16.13$, $p < .001$, no effect of test modality, $F(1,76) = 2.67$, $p = .11$ and no interaction, $F < 1$. The same 2 × 2 ANOVA on mean percent length revealed an effect of input modality and the same analysis for mean percent length showed an effect of input modality, $F(1,76) = 24.69$, $p < .001$, no effect of test modality, $F(1,76) = 2.65$, $p = .11$ and no interaction, $F < 1$.

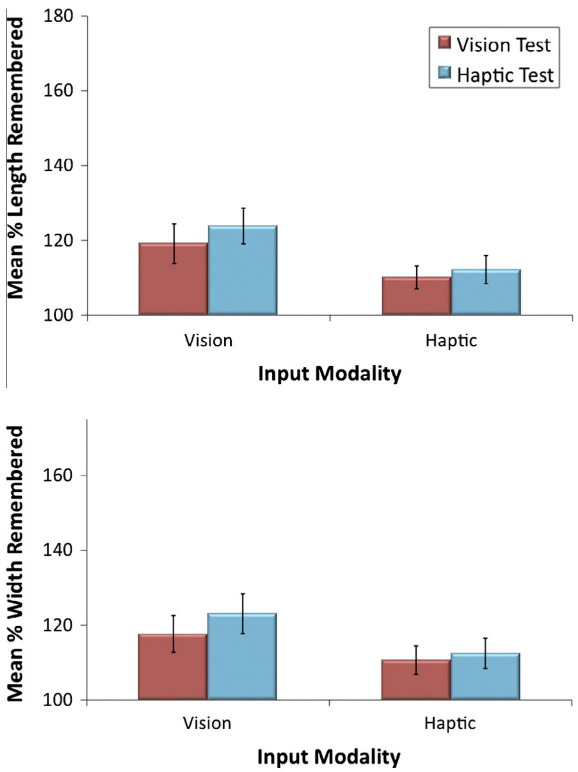

Fig. 5. Mean length (upper panel) and mean width (lower panel) remembered (%) as a function of the input modality and test modality (Experiment 1). Error bars indicate the 95% confidence interval for each mean.

2.2.3. Remembered boundary placement on each side

The mean percent remembered on the top, bottom, left and right sides for the vision input conditions was 118.8% (SD = 12.0), 123.9% (SD = 13.5), 119.6% (SD = 11.6) and 120.9% (SD = 13.4) respectively; and for the haptic conditions was 107.7% (SD = 8.1), 114.5% (SD = 9.7), 112.4% (SD = 10.4) and 110.6% (SD = 10.7) respectively. Single mean *t*-tests revealed that in each case the mean differed from 100% with $p < .001$; for the *vision condition*, $t(39) = 9.95$, $t(39) = 11.22$, $t(39) = 10.62$, $t(39) = 9.88$, respectively and for the *haptic condition*, $t(39) = 6.00$, $t(39) = 9.48$, $t(39) = 7.50$, $t(39) = 6.23$, respectively.

2.2.4. Memory for object placement

When asked whether they thought the experimenters had moved any of the objects with respect to one another, 65% of the participants in the vision exploration and 88% of participants in the manual exploration conditions stated that none of the objects had moved. Of the participants who said that an object had moved they tended to say it was a single object in 1–2 scenes. These single objects did not tend to be shifted in any one direction; participants reported an object being shifted inward, outward or as being in the same position but rotated.

2.2.5. Behavioral descriptions of visual and haptic exploration

The experimenter and assistant(s) observed performance closely. Vision participants honored the instructions not to look around the room, and to focus their gaze the designated scene region for the 30 s that their vision-blocking goggles were removed. Haptic participants followed instructions to use two hands in exploring the objects, but sometimes had to be reminded to do so. With reminders, all participants explored the regions up the boundaries. We noted that participants in the haptic condition tended to include "back and forth" hand movements between objects and between an object and the nearest boundary (or boundaries). Participants sometimes had to be reminded to feel the space up to the boundaries. But in all cases participants heeded the experimenter's reminders.

2.3. Discussion

Boundary extension occurred whether participants viewed the scene regions or explored them with their hands, but vision led to the more expansive boundary error. These observations held whether memory was tested using the same modality or the alternate modality. The majority of participants in all conditions correctly recognized that no objects were moved between study and test. Participants who reported a change, only referred to one or two objects (out of all the objects studied) and these false memories revealed no systematic displacement (such as a tendency to remember the objects as being shifted away from nearby boundaries; Huttenlocher, Hedges, &

Duncan, 1991). Participants simply remembered having perceived the continuation of the scene region beyond the stimulus' boundaries. Most important, the size of the boundary error associated with each modality transferred across modalities (without cost), suggesting that information had been retained in a functionally unitary scene representation (Kelly & Avraamides, 2011; Newell et al., 2005). In terms of the multisource model described earlier, the results suggest that there were not two independent multisource models each supporting a different amount of boundary extension, but that a single multisource representation supported memory for both modalities. To gain a better understanding of the nature of this multisource representation, in Experiment 2 we studied the effect of simultaneous bimodal exploration on boundary extension.

3. Experiment 2

In our interactions with the world we obtain information from more than a single modality at a time. In Experiment 2, we sought to determine the effect of bimodal exploration on memory for boundary placement. One possibility is that exploring the regions using simultaneous visual and haptic exploration might eliminate boundary extension. Participants would not only see and feel the objects and background of each scene-region, but would watch their own hands squeezing through the small spaces between the objects and boundaries, providing potentially salient adjunct information about boundary location. On the other hand, it is possible that participants would remember having seen and felt the world beyond the boundaries following bimodal exploration. As with one modality, the imagined continuation of the scene region for two modalities may simply be indistinguishable from the remembered sensory input, leading to boundary extension.

If bimodal boundary extension occurred, we planned to compare bimodal performance to each unimodal condition in Experiment 1 (*vision–vision* and *haptic–haptic*) to determine if bimodal boundary extension is the same as that observed for one modality alone. One possibility was that vision, the more "expansive" sense, would dominate. Another possibility was that the more "conservative" haptic sense might dominate. In either case this would suggest that information specific to the modality is retained with the amodal spatial framework. Results falling between the two would suggest an averaging of anticipated space.

3.1. Method

3.1.1. Participants

Participants were 20 (9 female) University of Delaware undergraduates, fulfilling a requirement for an introductory psychology course. No participants were excluded from the analysis.

3.1.2. Stimuli and apparatus

The stimuli were the same as in Experiment 1. Because vision was included in the bimodal condition, the visual frames were used during bimodal study and bimodal test. In this way, no visual information was occluded by the boundary, and because participants were viewing the displays, haptic movements were easily kept within the boundaries of the view.

3.1.3. Design and procedure

Design and procedure were the same as Experiment 1 except that participants used both vision and haptic exploration (simultaneously) during study and at test.

3.2. Results

All participants identified all objects and provided the same scene titles as in Experiment 1. The mean confidence rating in the bimodal condition was 2.1 ($SD = 0.3$); as in Experiment 1, mean confidence centered on "pretty sure (2)".

3.2.1. Remembered area

The mean area remembered following bimodal exploration is shown on the right side of the graph in Fig. 6; the confidence interval shows significant boundary extension. To provide a direct comparison of bimodal memory with the two unimodal conditions from Experiment 1, we present the mean area remembered in those conditions on the left side of the graph (Fig. 6).

A planned comparison analysis was conducted in which the *bimodal condition* was compared to each unimodal condition. The 3% difference in mean area increase between the bimodal and haptic conditions not approach significance, $t(57) = 0.53$, *n.s.*, $d = .2$. However the 18% difference between the bimodal and vision conditions was significant, $t(57) = 2.78$, $p < .01$, with Cohen's $d = .87$ (by convention, a large effect). Results showed that bimodal exploration led to no better memory for spatial expanse than did haptics alone. Thus, boundary extension appears to have been limited by the more conservative of the two modalities, haptics.[3]

The mean percent area remembered in each scene region is shown in Fig. 7. As can be seen in the figure, analysis of the individual scene-regions in the bimodal group shows that boundary extension occurred in memory for each of the seven scene regions.

Analysis of each participant's individual mean area remembered showed that of the 20 bimodal participants, only 1 mean was not greater than 100%.

3.2.2. Remembered length and width

Bimodal participants increased both the length, $M = 109.8\%$ ($SD = 6.5$), $t(19) = 6.70$, $p < .001$ and the width,

[3] Because this is a cross-experiment comparison, in considering the lack of a difference between the bimodal and haptic conditions, it is worthwhile to consider some additional descriptive information about the two distributions. They were very similar. In the bimodal condition, $SD = .14$ and Median = .20; in the haptic condition, $SD = .18$ and Median = .24. One participant in the haptic condition set an area increase greater than 2 SD above the mean and this score included an additional .34 above the next highest area increase in either of the two conditions. Removal of that participant reduced the 3% difference in mean area between the bimodal and haptic condition to 0% difference, with SDs of .14 and .13, respectively; and led to an identical range of scores. Thus, overall the distributions were very similar, and with removal of one participant with a particularly large area increase, were virtually identical.

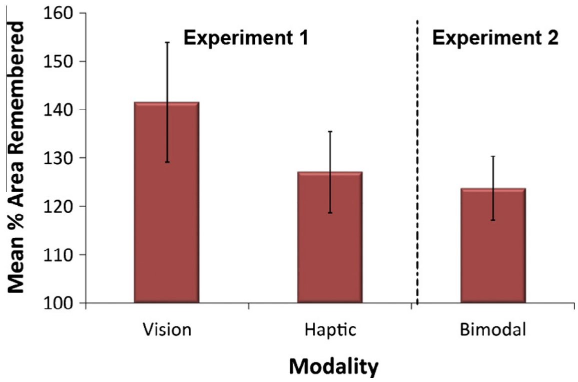

Fig. 6. Mean area remembered (%) in the vision–vision and haptic–haptic groups (Experiment 1) and the bimodal–bimodal group (Experiment 2). Error bars indicate the 95% confidence interval for each mean.

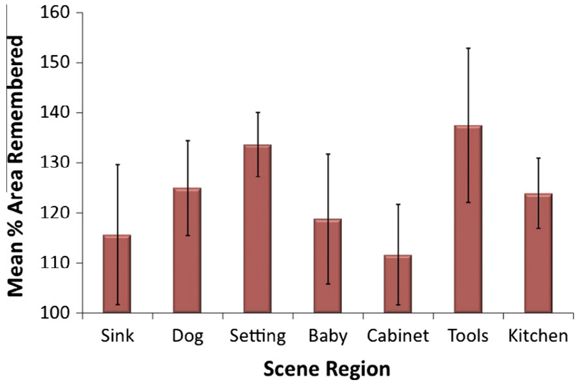

Fig. 7. Mean area remembered (%) in each scene-region in Experiment 2. Error bars indicate the 95% confidence interval for each mean.

$M = 112.2\%$ (SD = 7.0), $t(19) = 7.84$, $p < .001$ of the scene regions. To compare bimodal with unimodal performance (Experiment 1), we conducted the same planned comparison analysis we reported for changes in area, on changes in length and in width. Remembered increases in length and width did not differ between the bimodal and the haptic condition, $t(57) = 0.85$, $p = .40$ and $t(57) = 0.05$, $p = .96$, respectively; however vision was associated with a larger error in the length, $t(57) = 3.32$, $p < .01$, and an increase in the width that approached but did not reach significance, $t(57) = 1.90$, $p = .06$.

3.2.3. Remembered boundary placement on each side

Bimodal exploration led to significant boundary extension on the top, bottom, left and right sides of the regions 106.7% (SD = 7.6), 112.8% (SD = 7.2), 111.5% (SD = 8.1) and 112.9% (SD = 7.4), respectively. Single mean *t*-tests revealed that in each case the mean differed from 100% with $p < .01$; $t(19) = 3.94$, $t(19) = 7.92$, $t(19) = 6.37$, $t(19) = 7.78$, respectively.

3.2.4. Memory for object placement

When asked whether they thought the experimenters had moved any of the objects with respect to one another, 65% of the participants stated that none of the objects had moved. Of the participants who said that an object had moved they tended to say it was a single object in 1–2 scenes. Again, there was no directional bias: displacements were described as toward or away from the center or in the same location but rotated. A chi-square test was performed to examine the relation between reporting that an object had been moved and whether the input and test were bimodal or haptic (haptic–haptic condition in Experiment 1). The relation between these variables approached significance, X^2 (1, $N = 40$) = 3.58, $p = .06$.

3.3. Discussion

Observing one's own hands squeezing through the small spaces between the objects and boundaries during bimodal exploration did not eliminate boundary extension. Participants moved the boundaries out to reveal more than 24% more space than was in the stimulus region. Bimodal boundary extension did not differ from the 27% area increase in the haptic–haptic group from Experiment 1, but did significantly differ from the 42% increase set by the vision–vision group from Experiment 1. Bimodal boundary extension appeared to be constrained by the more "conservative" haptic modality (not the "dominant" visual modality). The similar performance between the bimodal and haptic groups may be been based on the presence of sensory qualia associated with haptics, or perhaps on a sense of space that was weighted fully to the haptic modality. In either case, there was no evidence of spatial "blending" across the two modalities. When bimodal participants attributed source at test, their boundary extension apparently reflected the more conservative modality.

4. Experiment 3

Participants in Experiments 1 and 2, and participants in Intraub (2004), all studied a small set of scene-regions (6–7 regions), and then received the boundary memory test minutes later. They knew that memory would be tested, but not the exact nature of the test. In addition, they navigated from one scene to next and moved through doorways (that they touched as part of the safety protocol of the experiment). In Experiment 3 we sought to determine if boundary extension in real space is similarly robust: (a) when the test instructions are presented in advance, prior to encoding (Gagnier et al., 2013; Intraub & Bodamer, 1993) and, (b) when memory is tested following each scene region (i.e., set size of one), as has been observed for multi-second picture presentations (i.e., set-size of one; Mullally et al., 2012; Seamon et al., 2002).

We wondered if in real-space, navigation from scene to scene between stimulus and test (walking along the same floor that supports all the scene regions) is necessary to support boundary extension. In addition, recent research has shown that movement through doorways between study and test can disrupt memory for objects (Radvansky, Krawietz, & Tamplin, 2011). For all of these reasons, it is important to determine if participants' memories for viewed scene regions would show boundary extension when participants do not move between study and test (as in picture memory studies) and when, they are provided with test information in advance, as has been demonstrated with pictures (e.g., Gagnier et al., 2013; Intraub & Bodamer, 1993).

Participants in Experiment 3 viewed the same seven scene-regions as in Experiments 1 and 2. Boundary memory was tested 30 s following study (the time required to remove the boundaries while the participant's vision was blocked with a curtain). Participants remained in the same position for study and test, and received all test instructions at the beginning of the experiment. Thus, they knew in advance that boundary memory would be tested.

4.1. Method

4.1.1. Participants

Participants taking part in the experiment were 20 (14 female) University of Delaware undergraduates, fulfilling a requirement for an introductory psychology course. Following our *a priori* rule for outliers described earlier (3SD or greater), one participant was excluded from the analysis, thus there were 19 participants (13 female) in the analysis. (We also analyzed the results including the outlier and there were no major differences; a brief description of those data is included in the results section.)

4.1.2. Stimuli and apparatus

The stimuli and apparatus were the same as Experiment 1.

4.1.3. Design and procedure

The design and procedure were the same as in the *vision–vision* condition of Experiment 1 with the following exceptions. After studying each scene for 30 s, a black cloth was raised in front of the participant to prevent him or her from seeing the scene-region. Participants then described the scene while the boundaries were removed by an assistant. After 30 s, the black cloth was lowered and participants performed the memory test. Participants were then blindfolded and led to the next scene, where the procedure continued until all scene-regions were tested.

4.2. Results

All objects were readily identified and titles to characterize the gist of the scene were similar to those in the previous experiments. Again, participants tended to report being "pretty sure" (2) of their reconstructions; the mean confidence rating was 2.2 (SD = 0.4).

4.2.1. Remembered area

On average, participants remembered having seen 115.1% of the area, this increase in area was significant $t(18) = 6.0$, $p < .01$. The mean percent area remembered for each scene is shown in Fig. 8; boundary extension occurred in memory for all seven scenes. The tendency to move the boundaries was evidence in 18 of the 19 participants; only one participant's mean area suggested a reduction in remembered area (i.e., a mean less than 100%).

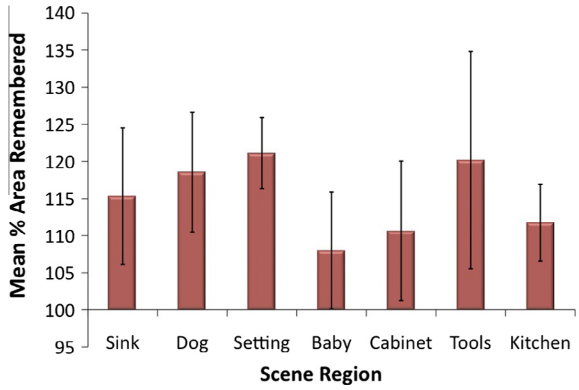

Fig. 8. The mean area remembered (%) for each scene-region in Experiment 3. Error bars indicate the 95% confidence interval for each mean.

4.2.2. Remembered length and width

Collapsing across scenes, the mean remembered width was 107.2% (SD = 5.6%) and mean remembered length was 107.2% (SD = 6.0%). Single mean *t*-tests comparing the remembered width and length to no change in length and width (100%) showed that the boundaries were moved outward to include more space than before, $t(18) = 5.53$, $p < .01$ and $t(18) = 5.22$, $p < .01$, respectively. The mean percent length and width remembered for each scene region is shown in Fig. 9. As shown in the figure, although length and width tended to shift outward, unlike the prior

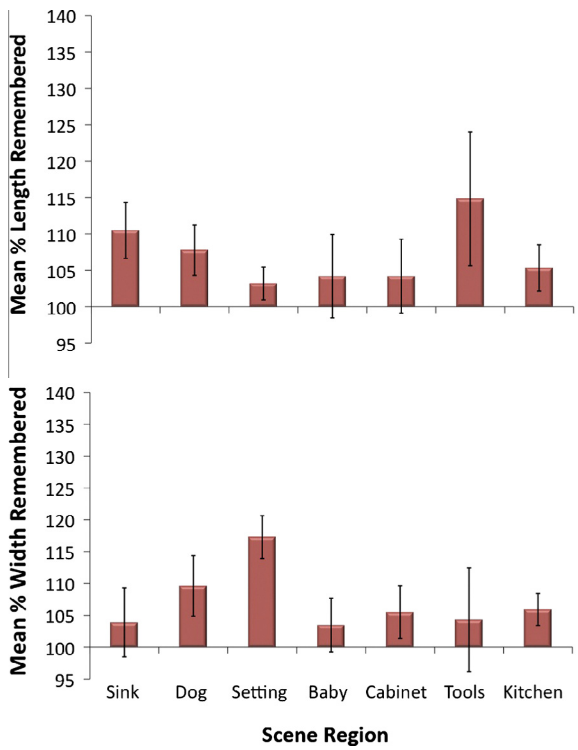

Fig. 9. The mean length (top panel) and width (bottom panel) remembered (%) for each scene-region in Experiment 3.

experiments, given the small overall boundary extension, not all such shifts reached significance.

4.2.3. Remembered boundary placement on each side

Overall, the mean remembered on the top, bottom, left and right sides was 102.8% (SD = 3.2), 105.9% (SD = 4.8), 103.1% (SD = 4.9) and 105.3% (SD = 4.6) respectively; each differed from 100%, $t(18) = 3.74$, $p < .01$, $t(18) = 5.34$, $p < .001$, $t(18) = 2.74$, $p < .02$ and $t(18) = 6.30$, $p < .001$, respectively.

4.2.4. Memory for object placement

When asked whether they thought the experimenters had moved any of the objects with respect to one another, 74% of the participants stated that none of the objects had moved. As in the previous experiments, when participants reported that an object had been moved they tended to say it was a single object in 1–2 scenes and there was no directional pattern in their description of what changed.

4.2.5. Analysis including the outlier

We also analyzed the data including the participant whose score fell beyond 3 SDs of the mean. Results were the same except that the mean area remembered for the *baby scene* approached but did not reach significance.

4.3. Discussion

Boundary extension occurred in memory for each of the scene-regions when participants remained in a single location (no navigation between stimulus and test), were forewarned that boundary memory would be tested, and memory was tested for each scene region only 30 s following study. Not surprisingly, inspection of the size of the boundary error in Experiment 3 compared to Experiments 1 and 2 suggests that these changes reduced boundary extension. Providing specific test information prior to study, for example, has been shown to attenuate the error (Gagnier et al., 2013; Intraub & Bodamer, 1993). What is important is that as in the case of pictures, boundary extension in real space was not eliminated under these conditions. Boundary extension in peripersonal space, like boundary extension in memory for photographs, appears to be part of scene representation that is rapidly present and not easily eliminated, even when participants know in advance exactly what will be tested.

5. General discussion

Boundary extension occurred in memory for bounded regions of peripersonal space following visual, haptic and bimodal study. The first question we addressed in this series of experiments was whether visual boundary extension and haptic boundary extension draw on a functionally unitary scene representation or on two separate modality-specific scene representations. Cross-modal tests in Experiment 1 suggested a common scene representation. Vision resulted in greater boundary extension than did haptics, replicating an earlier such observation with different scene-regions (Intraub, 2004), and this modality-related difference transferred in the cross-modal tests reported in Experiment 1. Vision led to greater boundary extension than haptics whether the test was conducted using vision, or the participant was blindfolded and explored the regions manually. In our tests, input modality, rather than test modality determined the size of the boundary extension error. Cross-modal transfer without cost is generally taken to suggest a functionally unitary representation (Kelly & Avraamides, 2011; Newell et al., 2005).

What is the nature of this unitary representation? The multisource model proposed by Intraub (2010, 2012) and Intraub and Dickinson (2008), provides one possible characterization. In the multisource model, the organizing structure of scene representation is the observer's sense of surrounding space. This spatial framework is best described by Tversky and colleagues from the perspective of the perceiver as the space, "in front of me", "behind me", "above me", "below me", and "to the right or left of me" (Bryant et al., 1992; Franklin & Tversky, 1990; Tversky, 2009). In the multisource model, this amodal spatial structure is the core of scene perception, organizing bottom-up sensory information and top-down-generated expectations about the likely surrounding scene that the stimulus only partially reveals. The unimodal and cross-modal results of Experiment 1 suggest that this framework can also support multiple sensory inputs (vision and haptics). In important ways, this spatial framework is similar to the "spatial image" described in Loomis et al.'s (2013) discussion of memory for multimodal stimuli in that it is a surrounding spatial structure (not limited to the frontal plane) and is an amodal structure. The difference is that the "spatial image" is instantiated in working memory; whereas the spatial structure in the multisource model is conceptualized as a "standing" mental structure that provides the basis of scene perception.

The second question we addressed was whether bimodal study (simultaneous visual and haptic exploration of the same scene-regions) would eliminate boundary extension (Experiment 2).. Bimodal study, not only provided participants with information from two modalities, but allowed them to view their own hands as they squeezed through some of the very small spaces between the objects and boundaries. However, bimodal study did not eliminate boundary extension. On average, participants increased the area by 24%. Furthermore, this did not differ significantly from the 27% area increase associated with haptics alone (*haptic–haptic* group; Experiment 1), whereas it differed significantly from the 42% increase in area set by participants who had studied the same regions using vision alone (*vision–vision* group; Experiment 1). Bimodal boundary extension was apparently constrained by the more "conservative", haptic sense. This can be thought of in two ways. Perhaps modality-specific information is maintained within the unitary scene representation. At test, participants in the bimodal condition, although they also viewed the scene, may have eschewed moving the borders as far out as participants in the visual condition because they did not remember having moved their hands that far from the objects. This is plausible given that participants spent considerable time exploring the seven regions and then returned to the same regions just minutes later. Another alternative is that participants may not have access to modality-specific aspects of their experience in memory, but that an amodal sense of studied space was weighted more heavily toward the more "conservative" sense (in this case "haptics"). Similar alternatives have been raised in research on visual and auditory spatial working memory; and it has been suggested that shorter retention intervals may be associated with memory for modal information with more long-term representations relying on an amodal code, but this has not yet been fully resolved (e.g., Giudice, Klatzky, & Loomis, 2009; Loomis, Klatzky, McHugh, & Giudice, 2012). It is difficult to make direct comparisons across these experiments and the current experiments given the differences in stimuli, task and modality, but clearly the same issues apply.

Finally, Experiment 3 demonstrated that like memory for photographs, when memory for each scene was tested before moving to the next, and participants were forewarned about the precise nature of the boundary reconstruction test, boundary extension still occurred. In this experiment, after visually studying a scene region for 30 s, following a 30-s retention interval, and standing in the same location, participants still moved the boundaries outward. This was interesting in that the objects and boundaries were in close physical proximity to the participant, and participants had the benefit of stereopsis, parallax and normal body sway during study – factors one might expect to help support a veridical memory of the spatial relationship of the objects to the boundaries. However, as is the case with pictures (see Hubbard et al., 2010; Intraub, 2010, 2012 for reviews), boundary extension in real space was surprisingly robust. It should be noted that boundary extension also occurred under the same set of conditions following haptic study of scene-regions (Mullally et al., 2012). Thus far, there is nothing to suggest that boundary extension in memory for regions of peripersonal space is fundamentally different than memory for views of the world shown in a photograph. It appears that regardless of whether study is visual or haptic, or the stimuli are 2D or 3D, when a view of the world is presented to a perceiver, the same anticipatory spatial error occurs. In the next section we will address a possible mechanism that may account for boundary extension under these various conditions.

6. Visual, haptic and bimodal boundary extension

In natural scene perception, the perceiver is typically embedded within the scene he or she perceives. For example, in contrast to viewing a picture of a kitchen, one stands *in* a kitchen with appliances, cabinetry and kitchen furniture surrounding the perceiver. Scenes don't exist solely in the frontal plane, as in a picture, but surround the perceiver. According to the multisource model, the mental representation of a scene is similarly structured in terms of surrounding space. The surrounding space just outside the boundary is most highly constrained by the information just inside the boundary (e.g., amodal continuation of the surface in combination with top-down knowledge), and the representation of anticipated space shades out to be increasing schematic and less well-defined the farther from the stimulus boundaries it is. Boundary extension is considered to be a source monitoring error, in which some of the most highly constrained anticipated spatial context just beyond the boundaries is attributed to perception. This is why boundary extension is limited to a small area just beyond the edges of the display; only information that most closely resembles remembered sensory information will be misattributed to having been perceived earlier.

In considering why vision (arguably the dominant sense for sighted individuals) led to greater boundary extension that haptics in Experiment 1 and in Intraub (2004), it is worthwhile to consider that although boundary extension is an error with respect to the stimulus, it is a good prediction about anticipated surrounding space. According to the source monitoring framework offered by Johnson et al. (1993), the greater the similarity in characteristics between information derived from difference sources (e.g., imagination and perception), the greater the likelihood of a source monitoring error. The modality-specific difference in boundary extension may be due to a number of causes. Perhaps the imagined[4] continuation of the scene beyond the boundaries is more similar to the remembered region within the boundaries when participants view the scene regions than when they explore them using only their hands. This might reflect a difference in the "density" of sensory information from *within* the boundaries that are provided by the two modalities. Haptics is a contact sense (touch), in which multiple high acuity regions (the fingertips) provide sensory detail, whereas in vision the high acuity region is limited to the fovea and there is a very large low-acuity periphery. Participants may therefore accept a greater swath of visually imagined space as having been seen before than in the case of haptics. The difference in boundary extension may also reflect the different "scope" of the two modalities. For example, a small eye movement can bring to view a larger new area of the world than can a small hand movement. Thus, decisions based on haptic input may be more conservative in nature. Our research cannot tease out what aspects of the two modalities support this difference. However, in closing, it is important to note that although the size of the anticipatory error differed between the two modalities, and is a sizeable error with respect to the stimulus area, relative to the surrounding world, in both cases, boundary extension itself is a small anticipatory error.

The notion that similarities between sources (perceived vs. imagined) in scene representation affect the size of boundary extension has received some support in experiments focusing on visual memory for pictures. For example, Intraub et al. (2008) superimposed numerals on briefly presented photographs. When participants undertook a demanding visual search task involving the numerals, thus dividing visual attention, memory for the pictures included a *greater* boundary-extension error than when the numerals were ignored. The authors suggested that without focal attention to the picture visual memory was

[4] We use the term "imagined" continuation to refer to the part of the representation that was not obtained through a sensory source, but that was derived instead from any or all of the top-down sources described earlier (amodal perception, knowledge about scenes, object-to-context associations).

compromised enough to increase the similarity between the remembered information and its imagined continuation. The alternative explanation, that dividing attention lead to greater anticipatory computation of surrounding space, is possible, but seemed less credible. Gagnier and Intraub (2012) contrasted spatial memory for the same scenes when they were presented in photographic form and as line drawings (created by tracing the photographs). Boundary extension was greater for the outline drawings. The suggestion here was not that outline sketches of scenes evoke a stronger prediction about surrounding space than do naturalistic photographs, but that the similarity between memory for lines in a line drawing, and the imagined continuation of those lines was greater than the similarity between memory for the multidimensional characteristics of the photographs and imagination of that more complex visual information. In all cases, boundary extension occurred, but a greater amount of boundary extension was observed under conditions in which similarity between the sensory input and the imagined extension of that input was arguably greater.

In Experiment 2, we had thought that seeing one's own hands squeezing through the small spaces between the boundary and objects and feeling the tight spaces simultaneously would eliminate boundary extension or reduce it dramatically. But it did not, nor was there an averaging between visual and haptic boundary extension. As discussed earlier, boundary extension was apparently grounded in the more conservative of the two modalities (in this case "haptics). The observation that participants were not swayed by the additional visual input, may help to explain a set of null effects observed in a series of visual–auditory bimodal boundary extension experiments conducted by Gagnier (2010) in her dissertation research. She selected sound effects that were rated as strongly evoking a sense of surrounding space when paired with particular photographs of scenes (e.g. a close-up of city traffic jam paired with city street sounds and honking cars). In six experiments, which included long durations (15 s visual presentations) as well as conditions intended to induce confusion errors (e.g., brief visual presentations, a 2-day retention interval, an incidental memory test), boundary extension was unaffected by the presence of these sound effects; performance did not differ when the same pictures were paired with silence (or unrelated guitar music). Here participants were grounded in the visual sense. The presence of the more expansive auditory information did not cause them to believe they had *seen* more of the world even though it felt it helped evoke a stronger sense of the surrounding space. Neither Experiment 2 nor Gagnier (2010) suggests that there are no conditions in which modalities might blend or become confused in scene memory. What those experiments show is that under conditions that typically give rise to boundary extension (e.g., a fairly small number of highly distinctive scenes), boundary extension clearly occurred, but was seemingly grounded in the more "conservative" modality. As an adaptive error, the more conservative bias may be worthwhile in preventing participants from falsely remember too great and area, but just enough to be most useful given the information at hand.

Finally, why then, given this "conservative" bent in how perceiver's judge remembered space, do they disregard what would seem to be useful information about boundary placement – such as the sight and feel of hands squeezing through a small space in these stimuli? The experimenter observed the participants repeatedly move their hands between the objects and nearly boundaries both at presentation and at test, yet they remembered having felt more of the background. It should be pointed out that the same "disregard" of apparently useful information has been reported in picture studies as well. Gagnier et al. (2013) studied this issue in a series of eye tracking experiments. They showed participants 12 photographs of single-object scenes (close-up views) for 15 s each; the main object was cropped by a picture boundary on half the trials. The cropped relation was thought to provide a good marker of boundary placement on the cropped side. Participants were told to remember the scenes in as much detail as possible and indeed, eye tracking revealed that on cropped trials, participants fixated the cropped region both at study and at test, yet when adjusting the boundaries (by moving the boundaries inward or outward to reveal more or less of the scene) they tended not only to complete the cropped object but to show more of the background beyond it on the previously cropped side. In another condition, participants were explicitly warned, *prior to study*, about the boundary adjustment task that would follow. In this case oculomotor activity increased in the cropped region, demonstrating that participants did consider this to be a valuable marker for the subsequent test. They fixated the cropped region sooner, more frequently and with longer fixations, yet again, boundary extension occurred on the cropped side at test, just minutes later. The bimodal results of Experiment 2 in the present study, and the results of Gagnier et al.'s (2013) eye tracking research, underscores the idea that the goal of the cognitive systems underlying scene perception isn't to retain a record of the spurious movements of the eyes and hands, but to understand the surrounding world – a world that can only be sampled a bit at a time. Part of this understanding of a scene region may be to imagine its likely surrounding context, a process that draws upon top-down expectations about the scene's continuity.

In the present research, in all three experiments, under very different types of study (visual, haptic or bimodal), the expectation of continuity can be inferred because participants claimed to have remembered experiencing the area just outside the boundaries of the stimulus. In all three cases, conceptual identification of the scene regions (a "kitchen scene", a "bathroom scene", a "workman's areas") was the same. Memory for object position (in an open-ended test) showed no difference in the types of spatial errors made with respect to object placement. When asked to report any changes in the objects' positions after viewing the scene at test, there was no evidence that participants remembered the objects as being closer together than before. Participants rarely erred regarding object placement, and when they did, they reported an object has having been shifted toward the center, toward the edges and in the same location but turned, equally often. This was of particular concern in evaluating boundary extension because participants were allowed to move the boundaries and not the objects. For example, had they remembered the objects as having been shifted inward, grouped more closely together toward the center of the original space (e.g., Huttenlocher et al., 1991), this would raise questions as to whether changes in boundary placement were actually reflecting boundary extension. However, results showed that participants didn't remember the objects as having been closer together. Participants did however, place the boundaries outward, creating a more expansive scene region.

7. Conclusion

After studying small regions of scenes in peripersonal space, participants remembered having explored beyond the boundaries of the view whether perception had been visual, haptic or both simultaneously. This suggests that irrespective of modality the representation of scene-regions is a *multisource* representation that reflects both bottom-up (sensory) and top-down sources of input. Our cross-modal and bimodal conditions suggest a functionally unitary representation in which specific characteristics associated with specific modalities may be retained (either as qualia or in terms of spatial weighting). This modality-specific information is sufficient to influence the amount of imagined space that will be misattributed to having been perceived (boundary extension). The imagined continuation of a studied scene region may play a fundamental role in allowing the perceiver to understand, anticipate and act upon a surrounding world that can only be sampled a part at a time; whether sampling is carried out by the eyes, the hands or both.

Acknowledgements

This research was supported in part by a grant from the National Institutes of Health (MH54688) to HI. The authors thank research assistants Terry Penn, Kyrianna Ruddy, and Caroline Murray for their valuable contributions in helping to run these experiments.

References

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

Avery, G. C., & Day, R. H. (1969). Basis of the horizontal–vertical illusion. *Journal of Experimental Psychology, 81*(2), 376–380.

Bar, M. (2004). Visual objects in context. *Nature Reviews: Neuroscience, 5*(8), 617–629. http://dx.doi.org/10.1038/nrn1476.

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.

Burgess, N. (2002). The hippocampus, space, and viewpoints in episodic memory. *The Quarterly Journal of Experimental Psychology, 55*(4), 1057–1080. http://dx.doi.org/10.1080/02724980244000224.

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. http://dx.doi.org/10.1016/j.cortex.2012.11.010, pii:S0010-9452(12)00329-2.

Day, R. H., & Avery, G. C. (1970). Absence of the horizontal–vertical illusion in haptic space. *Journal of Experimental Psychology, 83*(1), 172–173.

Dickinson, C. A., & Intraub, H. (2008). Transsaccadic representation of layout: What is the time course of boundary extension? *Journal of Experimental Psychology: Human Perception Performance, 34*(3), 543–555. http://dx.doi.org/10.1037/0096-1523.34.3.543, pii:2008-06370-003.

Diwadkar, V. A., & McNamara, T. P. (1997). Viewpoint dependence in scene recognition. *Psychological Science, 8*(4), 302–307. http://dx.doi.org/10.1111/j.1467-9280.1997.tb00442.x.

Epstein, R. A. (2008). Parahippocampal and retrosplenial contributions to human spatial navigation. *Trends in Cognitive Sciences, 12*(10), 388–396. http://dx.doi.org/10.1016/j.tics.2008.07.004, pii:S1364-6613(08)00199-X.

Epstein, R. A. (2011). Cognitive neuroscience: Scene layout from vision and touch. *Current Biology, 21*(11), R437–438. http://dx.doi.org/10.1016/j.cub.2011.04.037, pii:S0960-9822(11)00479-9.

Ernst, M. O., & Bülthoff, H. H. (2004). Merging the senses into a robust percept. *Trends in Cognitive Sciences, 8*(4), 162–169. http://dx.doi.org/10.1016/j.tics.2004.02.002, pii:S1364661304000385.

Fantoni, C., Hilger, J. D., Gerbino, W., & Kellman, P. J. (2008). Surface interpolation and 3D relatability. *Journal of Vision, 8*(7), 29.1–29.19. http://dx.doi.org/10.1167/8.7.29/8/7/29/ [pii].

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

Gagnier, K. M. (2010). *Rethinking boundary extension: The role of source monitoring in scene memory*. Retrieved from ProQuest Dissertations and Theses (Accession Order No. [854499379]).

Gagnier, K. M., Dickinson, C. A., & Intraub, H. (2013). Fixating picture boundaries does not eliminate boundary extension: Implications for scene representation. *The Quarterly Journal of Experimental Psychology, 66*(11), 2161–2186. http://dx.doi.org/10.1080/17470218.2013.775595.

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. http://dx.doi.org/10.1080/13506285.2012.703705.

Giudice, N. A., Betty, M. R., & Loomis, J. M. (2011). Functional equivalence of spatial images from touch and vision: Evidence from spatial updating in blind and sighted individuals. *Journal of Experimental Psychology: Learning, Memory & Cognition, 37*(3), 621. http://dx.doi.org/10.1037/a0022331.

Giudice, N. A., Klatzky, R. L., & Loomis, J. M. (2009). Evidence for amodal representations after bimodal learning: Integration of haptic–visual layouts into a common spatial image. *Spatial Cognition & Computation, 9*(4), 287–304.

Gottesman, C. V. (2011). Mental layout extrapolations prime spatial processing of scenes. *Journal of Experimental Psychology: Human Perception Performance, 37*(2), 382–395. http://dx.doi.org/10.1037/a0021434, pii:2011-06411-003.

Greene, M. R., & Oliva, A. (2009). Recognition of natural scenes from global properties: Seeing the forest without representing the trees. *Cognitive Psychology, 58*(2), 137–176. http://dx.doi.org/10.1016/j.cogpsych.2008.06.001, pii:S0010-0285(08)00045-5.

Hay, J. C., Pick, H. L., & Ikeda, K. (1965). Visual capture produced by prism spectacles. *Psychonomic Science, 2*(8), 215–216.

Hubbard, T. L., Hutchison, J. L., & Courtney, J. R. (2010). Boundary extension: Findings and theories. *The Quarterly Journal of Experimental Psychology, 63*(8), 1467–1494. http://dx.doi.org/10.1080/17470210903511236, pii:921748086.

Huttenlocher, J., Hedges, L. V., & Duncan, S. (1991). Categories and particulars: Prototype effects in estimating spatial location. *Psychological Review, 98*(3), 352. http://dx.doi.org/10.1037/0033-295X.98.3.352.

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

Intraub, H. (2004). Anticipatory spatial representation of 3D regions explored by sighted observers and a deaf-and-blind-observer. *Cognition, 94*(1), 19–37. http://dx.doi.org/10.1016/j.cognition.2003.10.013, pii:S0010027704000447.

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 Inc..

Intraub, H. (2012). Rethinking visual scene perception. *Wiley Interdisciplinary Reviews: Cognitive Science, 3*, 117–127. http://dx.doi.org/10.1002/wcs.149.

Intraub, H., & Bodamer, J. L. (1993). Boundary extension: Fundamental aspect of pictorial representation or encoding artifact? *Journal of Experimental Psychology: Learning, Memory & Cognition, 19*(6), 1387–1397.

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. http://dx.doi.org/10.3758/PP.70.7.1337, pii:70/7/1337.

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. http://dx.doi.org/10.1111/j.1467-9280.2008.02192.x, pii:PSCI2192.

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

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

Kelly, J. W., & Avraamides, M. N. (2011). Cross-sensory transfer of reference frames in spatial memory. *Cognition, 118*(3), 444–450. http://dx.doi.org/10.1016/j.cognition.2010.12.006, pii:S0010-0277(10)00298-2.

Kelly, J. W., Avraamides, M. N., & Giudice, N. A. (2011). Haptic experiences influence visually acquired memories: Reference frames during multimodal spatial learning. *Psychonomic Bulletin & Review, 18*(6), 1119–1125. http://dx.doi.org/10.3758/s13423-011-0162-1.

Klatzky, R. L., Lederman, S. J., & Reed, C. (1987). There's more to touch than meets the eye: The salience of object attributes for haptics with and without vision. *Journal of Experimental Psychology: General, 116*(4), 356–369. http://dx.doi.org/10.1037/0096-3445.116.4.356.

Loomis, J. M, Klatzky, R. L, & Giudice, N. A (2013). Representing 3D space in working memory: Spatial images from vision, hearing, touch, and language. In *Multisensory imagery* (pp. 131–155). New York: Springer. http://dx.doi.org/10.1007/978-1-4614-5879-1_8.

Loomis, J. M., Klatzky, R. L., McHugh, B., & Giudice, N. A. (2012). Spatial working memory for locations specified by vision and audition: Testing the amodality hypothesis. *Attention, Perception, & Psychophysics, 74*(6), 1260–1267. http://dx.doi.org/10.3758/s13414-012-0311-2.

Maguire, E. A., & Mullally, S. L. (2013). The hippocampus: A manifesto for change. *Journal of Experimental Psychology: General, 142*(4), 1180–1189. http://dx.doi.org/10.1037/a0033650.

Morein-Zamir, S., Soto-Faraco, S., & Kingstone, A. (2003). Auditory capture of vision: Examining temporal ventriloquism. *Cognitive Brain Research, 17*(1), 154–163, pii:S0926641003000892.

Mou, W., & McNamara, T. P. (2002). Intrinsic frames of reference in spatial memory. *Journal of Experimental Psychology: Learning, Memory & Cognition, 28*(1), 162–170.

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. http://dx.doi.org/10.1016/j.cub.2012.01.001, pii:S0960-9822(12)00002-4.

Newell, F. N., Woods, A. T., Mernagh, M., & Bülthoff, H. H. (2005). Visual, haptic and crossmodal recognition of scenes. *Experimental Brain Research, 161*(2), 233–242. http://dx.doi.org/10.1007/s00221-004-2067-y.

O'Keefe, J., & Nadel, L. (1978). *The hippocampus as a cognitive map*. Oxford: The Clarendon Press.

O'Regan, J. K. (1992). Solving the "real" mysteries of visual perception: The world as an outside memory. *Canadian Journal of 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. http://dx.doi.org/10.1016/j.neuron.2007.04.006, pii:S0896-6273(07)00256-5.

Pasqualotto, A., Finucane, C. M., & Newell, F. N. (2005). Visual and haptic representations of scenes are updated with observer movement. *Experimental Brain Research, 166*(3–4), 481–488. http://dx.doi.org/10.1007/s00221-005-2399-2.

Previc, F. H. (1998). The neuropsychology of 3-D space. *Psychological Bulletin, 124*(2), 123–164.

Radvansky, G. A., Krawietz, S. A., & Tamplin, A. K. (2011). Walking through doorways causes forgetting: Further explorations. *The Quarterly Journal of Experimental Psychology, 64*, 1632–1645. http://dx.doi.org/10.1080/17470218.2011.571267, pii:935037204.

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.

Shelton, A. L., & Mcnamara, T. P. (1997). Multiple views of spatial memory. *Psychonomic Bulletin & Review, 4*(1), 102–106. http://dx.doi.org/10.3758/BF03210780.

Simons, D. J., & Wang, R. F. (1998). Perceiving real-world viewpoint changes. *Psychological Science, 9*(4), 315–320. http://dx.doi.org/10.1111/1467-9280.00062.

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.

Wang, R. F., & Simons, D. J. (1999). Active and passive scene recognition across views. *Cognition, 70*(2), 191–210. http://dx.doi.org/10.1016/S0010-0277(99)00012-8.

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, pii:S09609822(11)00480-5.

Yamamoto, N., & Philbeck, J. W. (2013). Intrinsic frames of reference in haptic spatial learning. *Cognition, 129*(2), 447–456. http://dx.doi.org/10.1016/j.cognition.2013.08.011, pii:S0010-0277(13)00168-6.

Yamamoto, N., & Shelton, A. L. (2005). Visual and proprioceptive representations in spatial memory. *Memory & Cognition, 33*(1), 140–150. http://dx.doi.org/10.3758/BF03195304.