Systematic transition from boundary extension to contraction along an object-to-scene continuum

Jeongho Park

Department of Psychology,
Harvard University, Cambridge, MA, USA ✉

Emilie Josephs

Computer Science and Artificial Intelligence Laboratory,
Massachusetts Institute of Technology,
Cambridge, MA, USA ✉

Talia Konkle

Department of Psychology,
Harvard University, Cambridge, MA, USA ✉

After viewing a picture of an environment, our memory of it typically extends beyond what was presented, a phenomenon referred to as *boundary extension*. But, sometimes memory errors show the opposite pattern—*boundary contraction*—and the relationship between these phenomena is controversial. We constructed virtual three-dimensional environments and created a series of views at different distances, from object close-ups to wide-angle indoor views, and tested for memory errors along this object-to-scene continuum. Boundary extension was evident for close-scale views and transitioned parametrically to boundary contraction for far-scale views. However, this transition point was not tied to a specific position in the environment (e.g., the point of reachability). Instead, it tracked with judgments of the best-looking view of the environment, in both rich-object and low-object environments. We offer a dynamic-tension account, where competition between object-based and scene-based affordances determines whether a view will extend or contract in memory. This study demonstrates that boundary extension and boundary contraction are not two separate phenomena but rather two parts of a continuum, suggesting a common underlying mechanism. The transition point between the two is not fixed but depends on the observer's judgment of the best-looking view of the environment. These findings provide new insights into how we perceive and remember a view of environment.

Introduction

After viewing a photograph, our memory of it is not veridical—at times, memory for the scene generates content that extends beyond the edge of the original scene, and other times our memory of the view contracts from the edges. Since originally reported in Intraub and Richardson (1989), *boundary extension* and the related phenomenon of *boundary contraction* have been intensely studied to explore the many factors that contribute to the structure in these systematic memory errors (for reviews, see Hubbard, Hutchison, & Courtney, 2010; Intraub, 2010; Intraub, 2012). The dimension of *viewing distance* is a key factor underlying these memory errors: As the view increases in distance, boundary extension is reduced or even absent (Intraub, 2020; Intraub & Richardson, 1989). Further, sometimes boundary contraction is found, typically at the condition with the farthest viewing distance (Bertamini, Jones, Spooner, & Hecht, 2005; Chadwick, Mullally, & Maguire, 2013; Intraub, Bender, & Mangels, 1992; McDunn, Brown, Hale, & Siddiqui, 2016), and theories regarding the nature of boundary contraction and its relationship to boundary extension are still highly debated (see correspondence between Bainbridge & Baker, 2020b, and Intraub, 2020; Gandolfo, Nägele, & Peelen, 2023; Greene & Trivedi, 2022; Hafri, Wadhwa, & Bonner, 2022; Lin, Hafri, & Bonner, 2022).

Although the influence of viewing distance is evident, investigations into this factor have often been explored indirectly (e.g., cropped photographs to a few different degrees around a central object) or with only three to five sampled distances in a simulated environment (e.g., Bertamini et al., 2005), making it difficult to see a full pattern of how the memory errors change as a function of viewing distance. A recent study has provided some insight by testing memory errors in a large number of natural scene photographs (Bainbridge & Baker, 2020a). They found a strong correlation between the subjectively rated viewing distance of an image and

the direction of the memory errors, with a gradual change from boundary extension at close-scale views to boundary contraction at far-scale views. These transitions between memory errors as a function of viewing distance raise an intriguing question: What is the cognitive significance of the point where the direction of errors changes from boundary extension to contraction?

On one possible account, the point of memory transition may be related to whether the depicted space appears subjectively within reach versus subjectively out of reach. Previous work has shown that the distinction between peripersonal and extrapersonal space is very salient in visuospatial encoding (Cléry, Guipponi, Wardak, & Hamed, 2015), and neuropsychological studies have shown separable attentional mechanisms for near and far space (e.g., double dissociations in visual hemi-neglect) (Cowey, Small, & Ellis, 1994; Cowey, Small, & Ellis, 1998; Halligan & Marshall, 1991). Additionally, we have recently shown that the human visual system is sensitive to the distinction between views of intermediate-scale reach spaces and navigable-scale scenes (Josephs & Konkle, 2019; Josephs & Konkle, 2020). Thus, it is possible that the subjective reachability of an environment is a behaviorally relevant signal toward which perceptual representations are biased.

Another potential account of the transitional point is that it reflects a canonical view along the continuum—that is, a view between the extremes that has some perceptual and representational privilege. This hypothesis draws on object representation research, where objects have canonical orientations and visual sizes at which observers prefer to view them, and these settings serve as anchoring points in visual memory experiments (Blanz, Tarr, & Bülthoff, 1999; Konkle & Oliva, 2007; Konkle & Oliva, 2011; Palmer, 1981). Thus, it is possible that a similar phenomenon occurs in scene representation, where environments may be represented with respect to a canonical view. In fact, this account is consistent with the *memory schema hypothesis* (Intraub & Richardson, 1989; Intraub et al., 1992), which predicted boundary extension for closer views and boundary contraction for wide-angle views with respect to the "prototype" (though note that Intraub and colleagues have since updated their theory to a two-process *extension–normalization model*; for example, see Intraub et al., 1992; Intraub, Gottesman, Willey, & Zuk, 1996). Here, we explore the relationship between the canonical view of a scene and the transitions between extension and contraction errors, and revisit the theoretical links between these phenomena.

Thus, in the current studies we examine these two possible accounts for the point of transition in memory errors. To do so, we first built three-dimensional (3D) virtual environments (e.g., Bertamini et al., 2005; Park, Josephs, & Konkle, 2022), reflecting a variety of indoor categories with many different kinds of objects present, but critically with an identical layout structure of the walls, with a mid-height surface supporting a central object. Then, we finely sampled views of this space along a continuum, from a close-up view of the central object to a farther scene view of the entire indoor environment. Our experiments first mapped the direction of memory errors for views along this continuum across environments. Then, critically, we focused on understanding the transition between extension and contraction errors, relating it to independent judgments of both the subjective reachability of the view and the overall goodness of view.

Memory errors along an object–scene continuum

In the first experiment, we tested for memory errors for scene views taken from an object-to-scene continuum, using the same paradigm as Bainbridge and Baker (2020a). However, our stimuli were much more finely sampled along the viewing distance within tightly controlled virtual environments, allowing us to directly test how memory errors vary as a function of viewing distance.

Method

Participants

The Memory Experiment had 377 participants from Amazon Mechanical Turk (MTurk) (gender and age information was not collected). All participants gave informed consent and were compensated with $0.25 for each MTurk Human Intelligence Task (HIT). All procedures were approved by the Harvard University Human Subjects Institutional Review Board. The sample size for each condition of interest was 900 trials (45 per image × 20 environments); a bootstrap resampling method (Strong & Alvarez, 2019; see Supplementary Material) showed >95% power to detect the correlation between position and memory score, with 40 trials per condition, indicating this design is extremely high powered.

Stimuli

Computer-generated imagery (CGI) environments were constructed using the Unity video game engine (version 2017.3.0; Unity Technologies, San Francisco, CA). Twenty indoor environments were constructed, reflecting a variety of semantic categories (e.g., kitchens,
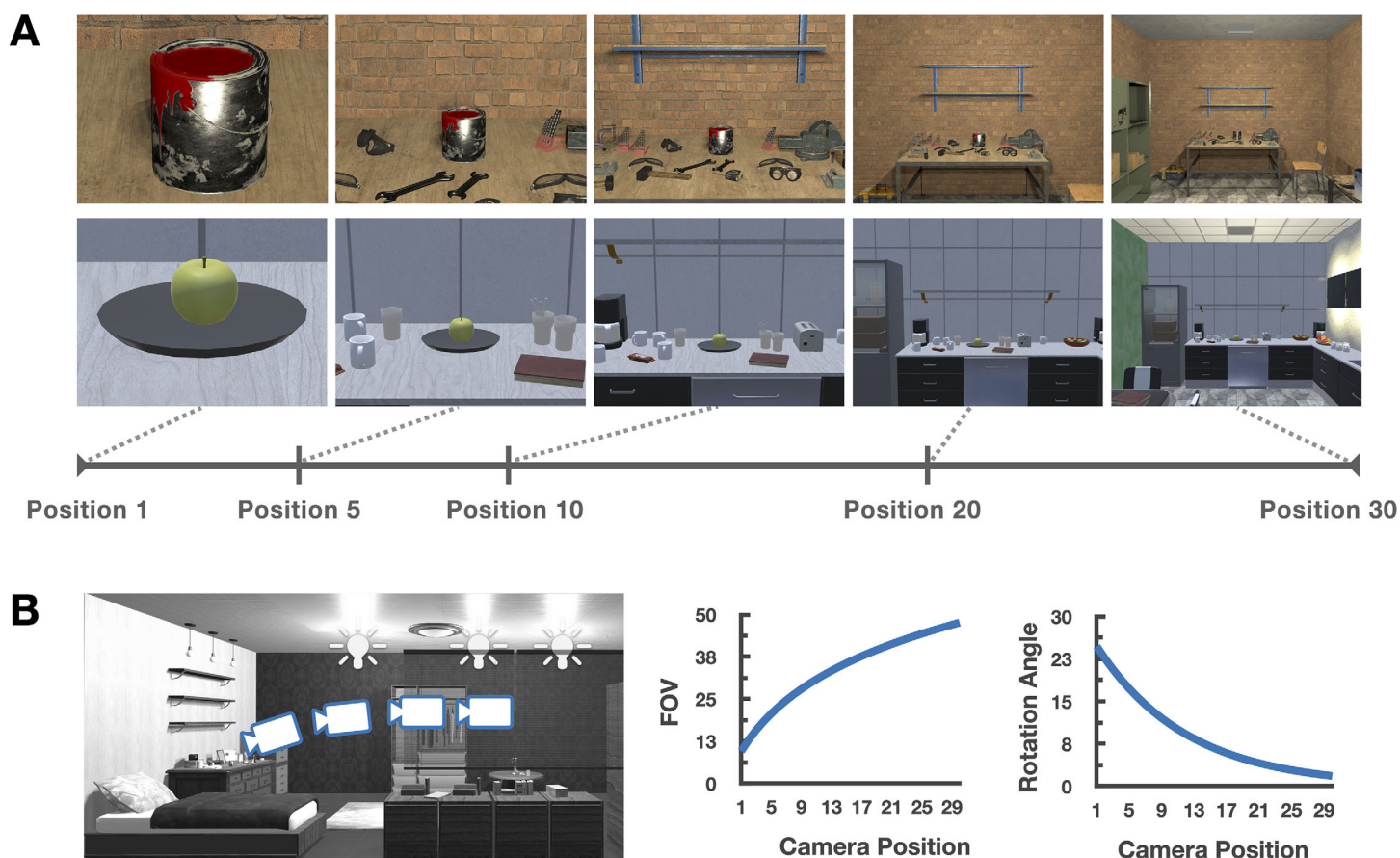
Figure 1. Stimuli. (**A**) Example images along the 30-point continuum from two different environments. (**B**) Schematic of the 3D environment in Unity with views taken from a camera from the front to the back of the scene. The FOV of the camera and the rotation angle of the camera were gradually changed to interpolate between an object-centered view of a single central object and a wider field of view of the entire scene.

bedrooms, laboratories, cafeterias). All rooms had the same physical dimensions (4 width × 3 height × 6 depth arbitrary units in Unity), with an extended horizontal surface along the back wall containing a centrally positioned object. Each environment was additionally populated with the kinds of objects typically encountered in those locations, creating naturalistic, "object-rich" CGI environments.

Images spanning a continuum of distances from the central object were captured from each environment, ranging from a close-up view of the object to a far-scale view that included the whole room. Images were generated by systematically varying the location of the camera (hereafter "Position") along 30 evenly spaced points arrayed from the "front" to the "back" of the room (i.e., from right in front of the central object to across the room from it) (Figure 1A). Close-up views were captured with a smaller camera field of view (FOV), so that only the central object appeared in the frame, and the FOV increased logarithmically with each step away from the object. The camera angle was parallel to the floor plane for far-scale views and was

gradually adjusted downward for closer positions, so that the central object was always at the center of the image (Figure 1B). These camera parameters were used for all 20 environments, yielding 600 unique stimuli (20 environments × 30 positions).

Experimental design

The experiment paradigm followed the procedures from Bainbridge and Baker (2020a). Each trial started with a central fixation cross for 1 second. The first stimulus was presented for 250 ms followed by five mosaic-scrambled mask images (50 ms each; random order). The second stimulus was then presented for 1 second. Unbeknownst to participants, the second stimulus was always identical to the first. After the second stimulus, participants were asked: "Compared to the first image, the second image is …," with two response options of "closer" or "farther" (Figure 2A). Participants were instructed to answer within 3 seconds using a keyboard; if no response was entered within this time, a warning message was shown before moving

A   Memory Paradigm
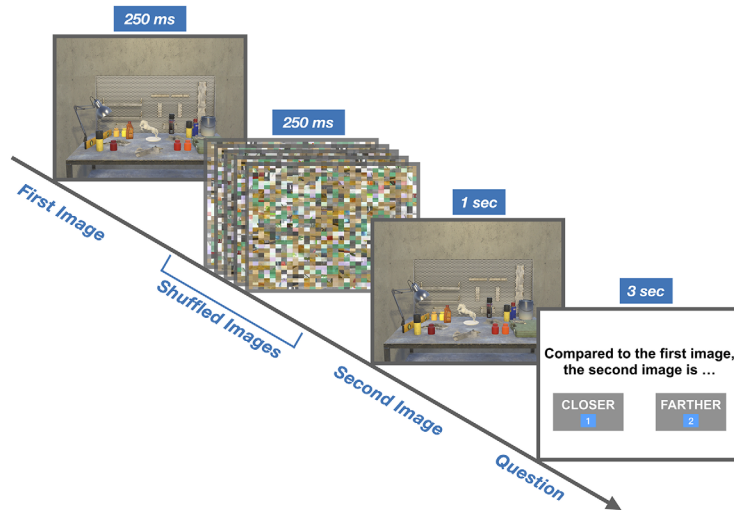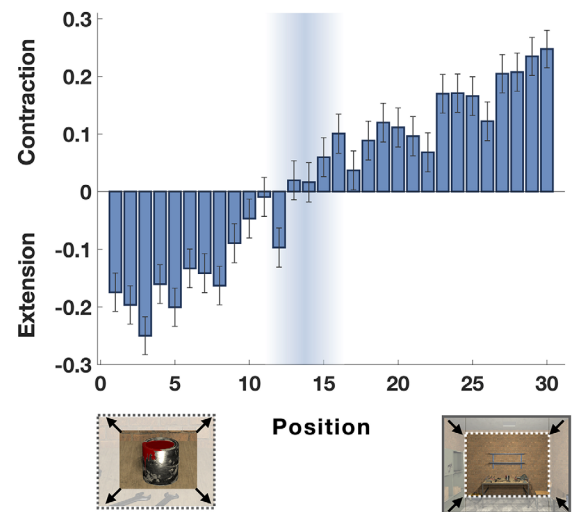

B   Memory Distortions

Figure 2. Memory Experiment paradigm and results. (**A**) Procedure of the Memory Experiment. The first image was shown for 250 ms followed by 250 ms of mosaic-scrambled masks. The second image was presented for 1 second, and participants were asked to answer whether the second image was closer or farther compared to the first image. Unbeknownst to participants, the second image was always identical to the first one. (**B**) Memory distortion scores averaged within each position. The error bars represent the standard error of the mean. The negative score indicates boundary extension, and the positive score indicates boundary contraction. Overall, there was a smooth transition from boundary extension to contraction as it changed from object-centered to scene-centered images.

to the next trial. Responses (closer/farther) and reaction times (RTs) were recorded.

Stimuli and masks were presented at 350 × 467 pixels. To create the masks, 15 CGI scenes (not used in the experiment) were broken down into 25 × 25 mosaic grids, then reassembled by randomly sampling from this pool.

The 600 images were divided into sets, such that a single HIT contained images from each of the 30 positions, selected from 10 of the 20 environments (thus, each HIT had three images from the same environment, viewed from different positions). The presentation order of stimuli within each set was randomly determined for each HIT. This method of dividing the images into sets ensured that all participants were exposed to the full range of views (i.e., Positions 1–30) within a HIT and allowed for the full stimulus set to be tested with 20 different image sets. Across all HITs and subjects, 45 trials were collected for each image.

Data preprocessing

Trials with reaction times exceeding the allowable response time window were excluded (3 seconds from stimulus onset; 2% of trials). HITs were excluded if (1) their average RT across all trials was faster than 3 *SD* from the mean over all participants, (2) more than 50% of their trials exceeded the allowable response time window, or (3) their average RT across all trials was faster than 2 *SD* from the mean and they answered more

than 90% of trials with the same response key, leading to the exclusion of 1.5% of the data. Reaction times were log transformed prior to trimming to account for the right skewness of RT distributions (Palmer, Horowitz, Torralba, & Wolfe, 2011; Ratcliff, 1979).

Calculating memory distortion scores

Responses of "closer" were assigned a score of −1, and "farther" responses were assigned a score of +1. These scores were averaged to obtain a *memory distortion score* for each image. Memory distortion scores thus fell in a range between −1 and 1, where negative scores indicate boundary extension (remembering the initial scene as further away than it was), positive scores indicate boundary contraction (remembering the scene closer than it was), and 0 indicates accurate position memory, with no distortion or bias one way or the other.

Linear mixed-effects logistic regression

Given the hypothesis that the size and direction of memory distortions would vary continuously as a factor of viewing distance, we modeled memory distortion scores as a factor of Position using linear mixed-effects logistic regression, implemented with the R code package lme4 (Bates, Ma¨chler, Bolker, & Walker, 2015). For the dichotomous response outcomes in this study (closer/farther), a binomial logit

link function was used in a generalized linear mixed model. The model included Position as a fixed-effect term and Environment and Subject as random intercepts.

To test the significance of the fixed effect of Position, we compared our model against a null model that had the same random effects structure but without the fixed effect, using a chi-square test. If the full model performed significantly better than the null model at predicting the data, then Position significantly contributed to the degree of memory distortion. The *p* values were estimated using the code package lmerTest with Satterthwaite approximation (Kuznetsova, Brockhoff, & Christensen, 2017).

Transition points estimation

If the direction of the memory distortion varied continuously as a function of viewing distance to the central object, then there should be a Position where there is no distortion or bias (i.e., average memory distortion score = 0). These *transition points* were estimated by finding the Position where the mean response score changed its sign from negative to positive. If more than one position crossed the mean score = 0, the transition point was determined as the average of those positions. We then estimated the *transition range* along the continuum, by computing a bootstrap 95% confidence interval (CI). Specifically, the data were resampled with replacement within each participant, and a transition point was estimated from the resampled data. These procedures were repeated 1000 times, resulting in a distribution of transition points. Then, the 95% CI was obtained by computing the 0.025 and 0.975 quantiles of this distribution.[1]

Transparency and openness

All stimuli and data are available in an Open Science Foundation repository (https://bit.ly/3o5L8zc). None of the studies was preregistered.

Results

Is there a consistent transition between extension and contraction in memory, as would be expected by Bainbridge and Baker (2020a)? Indeed, we found systematic evidence that this was the case (Figure 2B). That is, close-scale views elicited strong boundary extension effects (i.e., negative memory distortion scores; 44.8% of images), whereas far-scale views elicited strong boundary contraction effects (i.e., positive memory distortion scores; 51% of images), and 4.1% of the images showed no boundary effect. The magnitude of the memory bias (i.e., the proportion of responding

one answer to the other) varied continuously between these extremes (Spearman rank correlation between position and memory distortion: rho = 0.66, $p = 8.76 \times 10^{-77}$) (Supplementary Figure S3A), confirmed with mixed-effects logistic regression (estimated odds ratio = 1.03, $p < 0.001$).

Importantly, there was a relatively consistent transition range (Position 12.5; transition range = 11.2–16.2), depicted as a vertical gradient zone in Figure 2B, which reflects the bootstrap 95% CI. Given that our stimuli encompass a range of semantic scene categories with diverse objects, these data indicate that such systematic transition between memory extension and contraction likely depends on something that all scenes had in common: namely, the same virtual layout of walls, with a back counter or large object surface and a small central object (see Supplementary Figure S1 for estimated transition points for each environment).

Overall, these results confirm previous findings that the direction and magnitude of memory bias are related to the viewing distance of an image, and they demonstrate that there is a systematic point at which the memory distortion shifts from extension to contraction. We next explored two possible explanations of this transition range.

Perceived reachability and preferred views

Visual inspection of the images falling in the transition range reveals a set of views where the structures at the front of the room (i.e., tables or counters) appear just reachable with the hand. Thus, we reasoned that it was possible that the transition range in memory errors is linked to whether the depicted space is perceived as in or out of reach. We note that the "reachability" of an image is a fairly subjective property, linked to both the arm's length of the viewer and how they imagine situating themselves in the space depicted in a two-dimensional image. Thus, we asked an independent set of participants to rate the subjective reachability for every view along the continuum.

Another possibility is that the transition region reflects a canonical view (or prototype) to which the memory is biased. Here, we took the canonical view of the environment to be the view that "looks best," drawing on paradigms related to canonical orientation and visual size of objects (Blanz et al., 1999; Konkle & Oliva, 2007; Konkle & Oliva, 2011; Palmer, 1981), and related to theories of aesthetic preference (Palmer, Schloss, & Sammartino, 2013). Specifically, we obtained "goodness-of-view" judgments for every image from an independent set of participants.

Method

Participants

We recruited 213 participants for the reachability task and 257 participants for the goodness-of-view task (eight participants completed both tasks; 18 completed both reachability and the above memory distortion task; 23 completed both goodness-of-view and the memory distortion tasks). All participants were recruited on MTurk (gender and age information was not collected), gave informed consent, and were compensated with $0.25 per HIT. All procedures were approved by the Harvard University Human Subjects Institutional Review Board. The sample sizes of both tasks were matched to the Memory Experiment.

Experimental design

In each trial, an image (500 × 667 pixels) was displayed, and participants made an untimed judgment. For the reachability task, the instructions were to "Judge whether the view was within reach or out of reach." Participants chose among those two options in a forced-choice task. For the goodness-of-view task, the instructions were to "Judge how good the current view is. Let's say you are a photographer. Given this view, which would give you a better view?" Participants chose among three response options: taking one step forward, taking one step backward, and taking no steps. Forty-five ratings were obtained for each image, for both tasks. The same stimuli and counterbalancing method used in the Memory Experiment was used in both rating tasks. The presentation order of stimuli within each set was randomly determined for each HIT.

Data preprocessing

Any HITs with average RTs falling outside of 3 *SD* from the mean were removed. As above, all RTs were log transformed prior to trimming. This procedure was performed separately for the reachability dataset (1.3% excluded) and the goodness-of-view dataset (1% excluded).

Calculating reachability score and goodness-of-view score

For the reachability task, "out of reach" responses were assigned a score of +1 and "within reach" responses were assigned −1. These scores were averaged to obtain a *reachability score* for each image. Here, a negative score indicated a reachable view, and a positive score indicated a non-reachable view, in a range of [−1, 1]. For the goodness-of-view task, we assigned −1 for

"backward", 0 for "no move", and +1 for "forward," and we averaged the values for each image. A negative goodness-of-view score indicated a preference to move backward, and a positive score indicated a preference to move forward, within a range of [−1, 1].

Linear mixed-effects logistic regression

A linear mixed-effects logistic regression was used to fit the data from the reachability and goodness-of-view tasks, using lme4 (Bates et al., 2015) implemented in R (R Core Team, 2016). For the reachability dataset, the model included Position as a fixed-effect term and Environment and Subject as random intercepts and predicted a binary outcome (within reach or out of reach). For the goodness-of-view dataset, we excluded *no move* response trials for the logistic regression and modeled the data only with two responses (move forward or move backward); the model included Position as a fixed-effect term and Environment and Subject as random intercepts.

Comparing the transition points

A bootstrap CI was used to infer the statistical significance between the memory transition points and transition points of reachability or goodness-of-view judgments. First, a hypothetical dataset was created by sampling the data with replacement within each participant, independently for each experiment. Then, we computed transition points using the same methods as in the Memory Experiment and calculated the *difference* of transition points between the experiments. These procedures were repeated 1000 times, resulting in a distribution of differences in the transition points. The 95% CI was obtained by computing the 0.025 and 0.975 quantiles of this distribution. If this 95% CI of the difference contained 0 (i.e., no difference between the transition points), we inferred that there was no statistically significant difference ($p > 0.05$); whereas, if the 95% CI did not contain 0, then we inferred that there was a statistically significant difference ($p < 0.05$).

Results

Do either the reachability or goodness-of-view judgments have transition points in a similar range as the Memory Experiment? First, we assessed reachability ratings, which showed a significant effect of Position (estimated odds ratio = 1.52, $p < 0.001$) and a transition point near the front of the room, as expected (Position 10.5; transition range = 9.5–10.5) (Figure 3B). It is noticeable that the reachability ratings had a much smaller variance compared to the Memory Experiment, suggesting that there was strong consensus
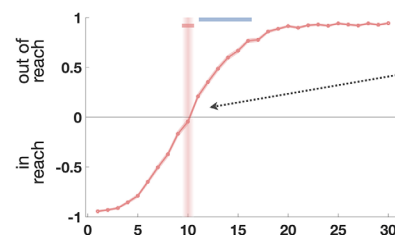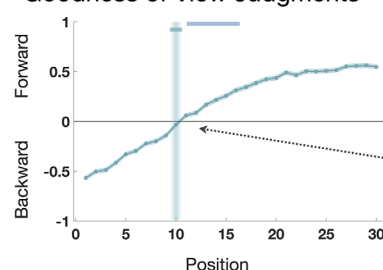
Figure 3. Transition range examples and results from reachability and goodness-of-view judgments. (**A**) Examples of the memory transition points from four different environments. (**B**) The reachability judgment score averaged by each position and an example image of the transition point. The negative score represents within-reach judgments, the positive represents out-of-reach judgments, and the zero represents the transition point of subjective reachability. (**C**) The goodness-of-view judgment score averaged by each position and an example image of the transition point. The negative score represents a preference to step backward, the positive represents a preference to step forward, and the zero represents the most preferred ("looks good") view. For both (**B**) and (**C**), a vertical gradient bar indicates a transitional range (bootstrap 95% CI), and the horizontal bar on top of the plot shows the transition range from memory distortions (light blue).

across different individuals and stimulus environments regarding whether the current view was reachable or not. This reachability transition range was very similar to the boundary extension/boundary contraction transition (visually indicated in Figure 3B with the horizontal bars for reference), although significantly closer to the front of the room (difference = 2.0; 95% CI, 0.7–6.2; $p < 0.05$). However, this significant difference was quite small, as perceptually evident when inspecting the example images shown in Figure 3B. Thus, the direction of memory distortions might indeed be linked to whether the space is perceived to be within or out of reach.

Considering the goodness-of-view ratings, we also found a smooth transition along the continuum (i.e., significant effect of Position, with estimated odds ratio = 1.14; $p < 0.001$), where participants chose to step backward for near-scale images but chose to step forward for far-scale images (Figure 3C). Interestingly, the transition point between these response options (i.e., the preferred or canonical view) was also in the same range (Position 10.5; transition range = 9.5–10.5) with the reachability transition (difference = 0; 95% CI, −1 to 1; $p < 0.05$). Additionally, the "no move" responses peaked around Positions 10 and 11 (Supplementary Figure S4A), further supporting the existence of a

consistent canonical view in this region along the continuum. Thus, the memory distortion effect might also be anchored around a canonical view of the environment.

Taken together, from the first set of experiments, we found suggestive evidence that both reachability and the goodness-of-view transition might track with the transition in memory errors. Is it possible to modify the virtual environments to dissociate reachability and goodness-of-view judgments?

We reasoned that reachability judgments should be relatively insensitive to the number and position of the surrounding objects in the environment, as such judgments are a factor of the length of the viewer's arm and the presence of a target relative to which they can measure distance. In contrast, goodness-of-view judgments may be relatively more sensitive to the overall distribution of multiple objects in the image (Leyssen, Linsen, Sammartino, & Palmer, 2012). Thus, in the following experiments, we created "low-object environments," predicting that these changes in the object content of environments would allow us to further dissociate reachability and goodness-of-view judgments.

Further, many studies have shown that the direction and strength of memory errors are influenced by the object content of a scene, including the retinal size of

the main object (Bertamini et al., 2005), the salience of the main object (Gallagher, Balas, Matheny, & Sinha, 2005), and the level of clutter around the main object (Gottesman, 2012; Gottesman, 2011; Gottesman, 2014). Based on this prior literature, it is likely that the transition range in memory errors will also change in these low-object environments, relative to the rich object environments used in the first set of experiments. Thus, in a second set of experiments, we measured memory errors, reachability, and goodness of view in low-object environments.

Low-object environments

A second stimulus set of visual environments was created from the original environments, but with all of the small and manipulable objects removed (e.g., cups, books), leaving only the immovable surfaces and furniture-like structures (e.g., kitchen countertops, large tables) (Figure 4). Thus, these low-object environments shared the same spatial layout and background content as the original object-rich environments but differed substantially in their object content. Next, images along the continuum were generated using the same parameters as the previous experiments, resulting in 600 unique stimuli (20 environments × 30 positions).

We completed the same three experiments reported above but used the new low-object stimulus set. All experimental details were identical. Data were collected from 173 participants in the Memory Experiment, 213 participants in the reachability task, and 153 participants in the goodness-of-view task (gender and age information was not collected). The data trimming procedures led to the exclusion of 2.8% of trials from the Memory Experiment dataset, 1.5% of trials from

the reachability dataset, and 0.7% of trials from the goodness-of-view dataset. As these paradigms were identical to the first set of experiments, the sample size was determined based on a power analysis on the rich-object environments data. The results are shown in Figure 5A.

In these low-object environments, the reachability and goodness-of-view transition ranges were quite different. Specifically, as expected, the reachability ratings remained at the front of the environment (Position 13.5; transition range = 12.5–13.5) (Figure 5A, bottom row), comparable between the low-object and rich-object environments. However, the goodness-of-view ratings were shifted toward a much farther view of the scene (Position 21.5; transition range = 20.5–22.5) (Figure 5A, middle row), showing a much larger transition point change between low-object and rich-object environments for the goodness of view than reachability. Example views of the reachability and the goodness-of-view transition are shown in Figure 5B.

Critically, the transition point in memory distortions was also shifted substantially further back (Position 23.5; transition range = 18.1–27.2) (Figure 5A, top row; see Supplementary Figure S2 for estimated transition points for each environment). The memory transition range tracked closely with the goodness-of-view transition range (difference = 2; 95% CI, −3.4 to 5.7; $p > 0.05$) and was significantly different from the reachability range (difference = 10; 95% CI, 4.6–14; $p < 0.05$). Thus, these experiments clearly demonstrate that that transition range in memory is not linked to the point of reachability of the view but instead seems to be linked to properties of the image or environment that also manifest in preference judgments.

Finally, one possible critique of the paradigm used here to evaluate boundary extension is that it
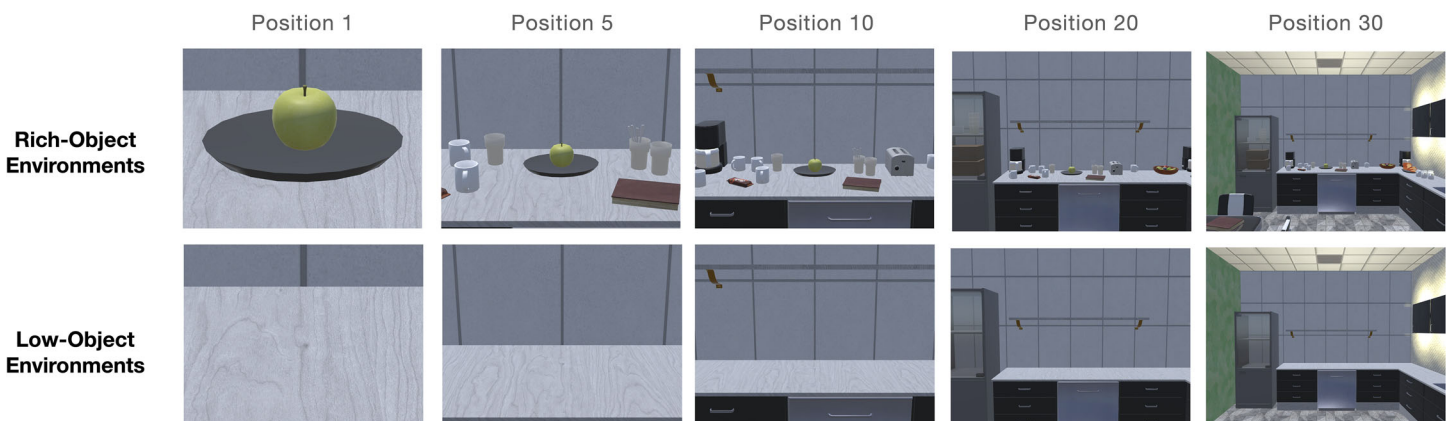

Figure 4. Low-object environments. All of the small and manipulable objects were removed from the rich-object environments, leaving the immovable surfaces and furniture-like structure present such as kitchen countertops or cabinets. As a result, two sets of environments shared the same spatial layout but differed substantially in their object content.
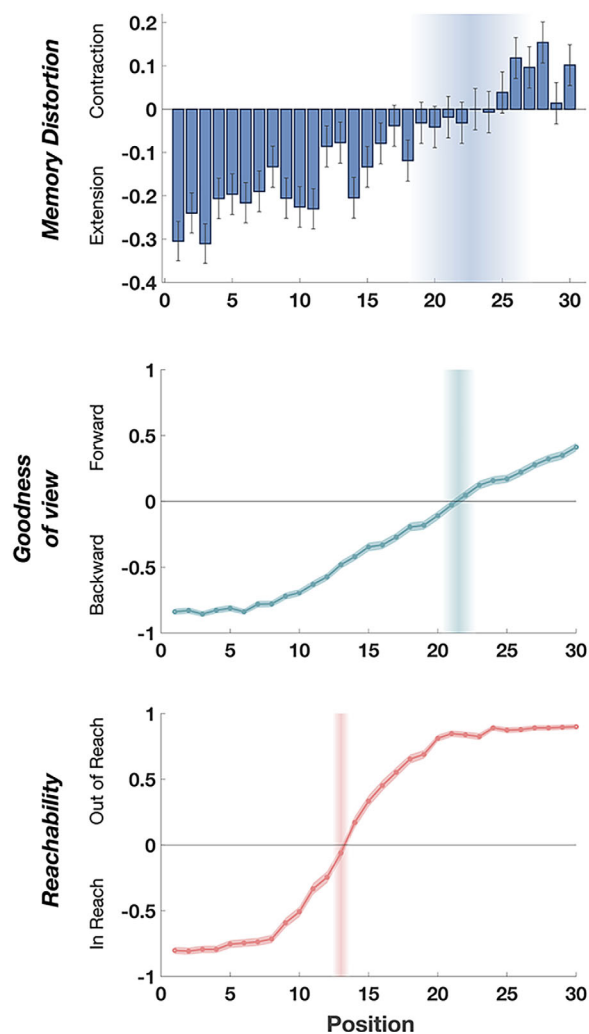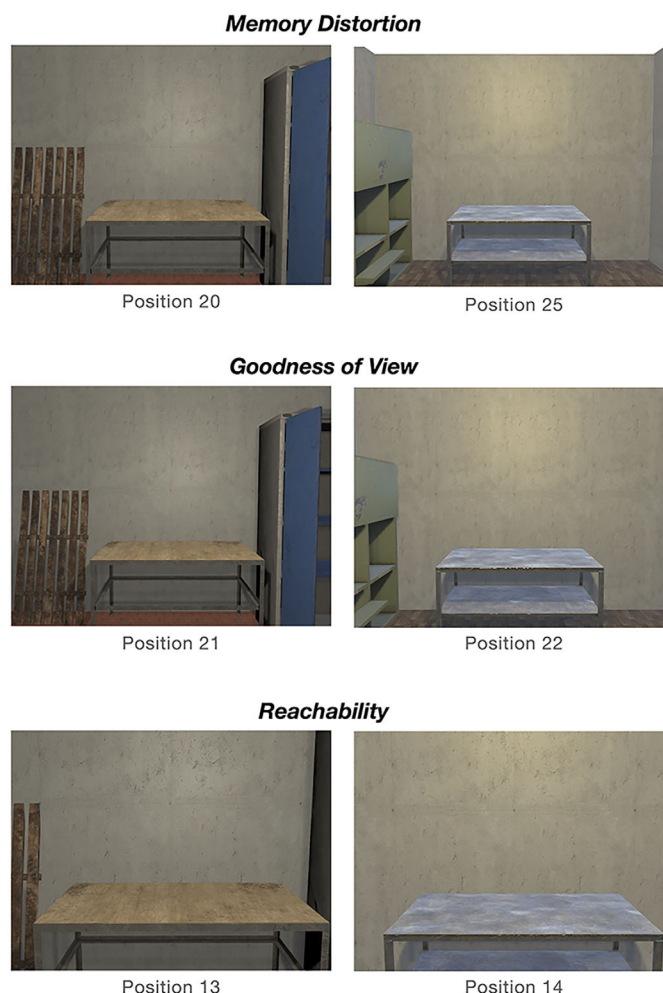
A Results (Low Object Environments)

B Transition Point Views

Figure 5. Low-object environment results and transition points. (**A**) The mean score by position from each paradigm. Positions around zero (*y*-axis) indicate the transition point, and a vertical gradient bar represents the transitional region (bootstrap 95% CI). Compared to the rich-object environments (Figures 2 and 3), the transition region shifted much further in the memory distortion and goodness-of-view judgments. However, the transition region of subjective reachability remained in relatively similar positions regardless of the stimulus set. (**B**) Examples of transition points from each paradigm. Critically, the transition range of goodness of view showed much closer resemblance to that of memory distortion, favoring the canonical view account.

is the kind that is likely to induce "normalization" based on the range of viewing distances present in the stimulus set. However, in this second set of experiments, we have the same range of viewing distances but a different transition point. Thus, these two experiments together indicate that a simple generic distance-normalizing mechanism (e.g., to the center of the environment) cannot fully account for these data. Instead, these data highlight that the transition between boundary extension and contraction depends on the interplay between object and scene processing, and they contribute new empirical evidence that this same interplay is linked to the preferred view of the environment.

General discussion

The goal of the current study was to map the transition point between boundary extension and boundary contraction memory errors along an object-to-scene continuum and to consider two hypotheses about the location of this transition point. Our approach leveraged custom-made 3D environments and systematically sampled views in these spaces, covering a more extensive range than previous studies. We found consistent transitions between boundary extension and boundary contraction, where the likelihood of the memory distortion varied highly systematically

with the viewing distance of the image, replicated in both object-rich and low-object environments. And, we found that the point in the environment at which there was no memory distortion was not linked to a specific viewing distance in the scene, such as the point of reachability, but was instead linked to judgments of the view that "looked best." Broadly, these results highlight that insight into the systematic distortions in scene memory can be gained by understanding the relationships among boundary extension, contraction, and the canonical view of the environment.

One possible relationship is a *prior-based* account of these phenomena, which draws on the reconstructive memory literature; that is, perceptual traces are encoded with respect to a structured prior distribution (e.g., Bartlett, 1932; Hemmer & Steyvers, 2009b; Huttenlocher, Hedges, & Vevea, 2000). In this Bayesian framework, systematic memory biases are signatures of optimal encoding of noisy perceptual input (e.g., Hemmer & Steyvers, 2009a). On such an account, each view of a scene is encoded with respect to some scene prior, which presumably includes information about a canonical vantage point within the space, biasing memory toward that point and yielding parametric extension or contraction. Because judgments of the best-looking views are also thought to draw on this internal scene prior (Gardner & Palmer, 2010), this account can also explain why the goodness-of-view ratings and memory errors closely tracked each other. However, this account alone does not articulate what scene properties are relevant for the prior; for example, it does not make a priori predictions that the canonical view would shift back in low-object environments. Thus, this account requires further specification; currently it can likely account for any pattern of data without making specific predictions for the direction of memory errors under different environmental configurations.

A more perceptual, *dynamic-tension* account of these data—one that we favor—is that, for any view, a balance of two oppositional forces determines whether subsequent memory will be extended or contracted. Each view is encoded in parallel with scene-based and object-based mechanisms. The scene-based mechanism is sensitive to global scene properties, such as concavity, navigability, openness, and mean depth (Bonner & Epstein, 2017; Cheng, Walther, Park, & Dilks, 2021; Greene & Oliva, 2009; Park & Park, 2020), and is biased to construct a broader view, perhaps via the amodal scene construction originally proposed by Intraub (2012). At the same time, the object-based mechanism exerts pressure to process objects in more detail or with higher fidelity, perhaps via object-based attention (Beighley, Sacco, Bauer, Hayes, & Intraub, 2019), resulting in down-weighting or loss of peripheral information.

In this dynamic-tension proposal, the view that "looks best" is the one that balances these opposing

affordances (rather than corresponding to the peak of an internal distribution). This account also naturally lends an explanation to some key patterns of our data. Rich-object environments, with their prominent object content, may have closer transition points (and preferred views) to balance the needs of object recognition, whereas low-object environments have farther transition points to favor spatial analysis of the environment. Indeed, the view that looks best for the low-object environments is just about where you start to see the side walls (Figure 5), which provide a stronger sense of 3D boundary and concavity to a viewer. It is an open question if the semantic relatedness of the object and scene mechanisms matter; for example, we predict that the object-processing component is relatively general and would operate similarly over novel or semantically incongruent objects in the scene of matched physical size. Overall, this account generally promotes thinking of the interplay between competing object-processing and spatial-layout processing mechanisms (Mullin & Steeves, 2013), such as predicting sensitivity to attentional manipulations and task demands, beyond static image-based information.

Finally, it is important to note that, in order to obtain a smooth continuum between scene-centered and object-centered views, we varied the FOV and distance, coming into a close crop of the object for the closest views to be compatible with previous studies. However, during naturalistic viewing, the vergence of the eyes on an object changes the human FOV far less dramatically than implied here, and visual content is not cut off with an image border. Future research could explore memory distortions in full-field viewing (e.g., Park, Soucy, Segawa, Mair, & Konkle, 2023), allowing far-peripheral visual processing to be engaged in situating the content and operations of central vision. Given the importance of central and peripheral information for visual system organization (Arcaro & Livingstone, 2017; Hasson, Harel, Levy, & Malach, 2003; Knapen, 2021), this may be a productive line of inquiry, moving the study of visual scene memory beyond the picture plane into agent-directed movements through an environment.

Keywords: scene memory, canonical views, virtual environments, aesthetic preferences

Acknowledgments

manuscript is posted on the PsyArXiv preprints server (https://psyarxiv.com/84exs).

Commercial relationships: none.
Corresponding author: Jeongho Park.
Email: jpark3@g.harvard.edu.
Address: Department of Psychology, Harvard University, Cambridge, MA, USA.

Footnote

[1]Note that, in an earlier version of this manuscript, we used a logistic function equation and fitted coefficients to estimate the transition, but in revision we found that the single-subject data were poorly fitted with this method. The current manuscript describes a more data-focused method we adopted to estimate the transition point. To further check for the stability of this procedure for estimating the transition points, we explored two types of variations. First, we estimated transition points with different levels of moving-average smoothing. For this, the mean scores were first smoothed across positions by computing the rolling average with a given window size, which was varied from two to six positions. We found that across these smoothing procedures the estimated transition points varied only minimally, approximately one Position. Second, we applied different constraints on how the data are resampled in the bootstrap procedure. Whether resampling within each position (across participants) or within each position and environment, we found similar CIs for the position range, with no qualitative differences in the overall patterns of results.

References

Arcaro, M. J., & Livingstone, M. S. (2017). Retinotopic organization of scene areas in macaque inferior temporal cortex. *Journal of Neuroscience, 37*(31), 7373–7389.

Bainbridge, W. A., & Baker, C. I. (2020a). Boundaries extend and contract in scene memory depending on image properties. *Current Biology, 30*(3), 537–543.

Bainbridge, W. A., & Baker, C. I. (2020b). Reply to Intraub. *Current Biology, 30*(24), R1465–R1466.

Bartlett, F. (1932). *Remembering: A study in experimental and social psychology*. Cambridge, UK: Cambridge University Press.

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48.

Beighley, S., Sacco, G. R., Bauer, L., Hayes, A. M., & Intraub, H. (2019). Remembering: Does the emotional content of a photograph affect boundary extension?, *Emotion, 19*(4), 699.

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.

Blanz, V., Tarr, M. J., & Bülthoff, H. H. (1999). What object attributes determine canonical views? *Perception, 28*(5), 575–599.

Bonner, M. F., & Epstein, R. A. (2017). Coding of navigational affordances in the human visual system. *Proceedings of the National Academy of Sciences, USA, 114*(18), 4793–4798.

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.

Cheng, A., Walther, D. B., Park, S., & Dilks, D. D. (2021). Concavity as a diagnostic feature of visual scenes. *NeuroImage, 232*, 117920.

Cléry, J., Guipponi, O., Wardak, C., & Hamed, S. B. (2015). Neuronal bases of peripersonal and extrapersonal spaces, their plasticity and their dynamics: Knowns and unknowns. *Neuropsychologia, 70*, 313–326.

Cowey, A., Small, M., & Ellis, S. (1994). Left visuo-spatial neglect can be worse in far than in near space. *Neuropsychologia, 32*(9), 1059–1066.

Cowey, A., Small, M., & Ellis, S. (1998). No abrupt change in visual hemineglect from near to far space. *Neuropsychologia, 37*(1), 1–6.

Gallagher, K., Balas, B., Matheny, J., & Sinha, P. (2005). The effects of scene category and content on boundary extension. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 27, 744–749.

Gandolfo, M., Nägele, H., & Peelen, M. V. (2023). Predictive processing of scene layout depends on naturalistic depth of field. *Psychological Science, 34*(3), 394–405.

Gardner, J. S., & Palmer, S. E. (2010). Representational fit in position and perspective: A unified aesthetic account. *Journal of Vision, 10*(7), 1232, https://doi.org/10.1167/10.7.1232.

Gottesman, C. (2011). More space please! The effect of clutter on boundary extension. *Journal of Vision, 11*(11), 1130, https://doi.org/10.1167/11.11.1130.

Gottesman, C. (2012). Effects of clutter on boundary extension: Volume or detail effects? *Journal of Vision, 12*(9), 1073, https://doi.org/10.1167/12.9.1073.

Gottesman, C. (2014). Disambiguating the effect of clutter on boundary extension. *Journal of Vision, 14*(10), 871, https://doi.org/10.1167/14.10.871.

Greene, M., & Trivedi, D. (2022). Spatial scene memories contain a fixed amount of semantic information. *PsyArxiv*, https://doi.org/10.31234/osf.io/r5fn9.

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.

Hafri, A., Wadhwa, S., & Bonner, M. F. (2022). Perceived distance alters memory for scene boundaries. *Psychological Science, 33*(12), 2040–2058

Halligan, P. W., & Marshall, J. C. (1991). Left neglect for near but not far space in man. *Nature, 350*(6318), 498–500.

Hasson, U., Harel, M., Levy, I., & Malach, R. (2003). Large-scale mirror-symmetry organization of human occipito-temporal object areas. *Neuron, 37*(6), 1027–1041.

Hemmer, P., & Steyvers, M. (2009a). A Bayesian account of reconstructive memory. *Topics in Cognitive Science, 1*(1), 189–202.

Hemmer, P., & Steyvers, M. (2009b). Integrating episodic memories and prior knowledge at multiple levels of abstraction. *Psychonomic Bulletin & Review, 16*(1), 80–87.

Hubbard, T. L., Hutchison, J. L., & Courtney, J. R. (2010). Boundary extension: Findings and theories. *Quarterly Journal of Experimental Psychology, 63*(8), 1467–1494.

Huttenlocher, J., Hedges, L. V., & Vevea, J. L. (2000). Why do categories affect stimulus judgment? *Journal of Experimental Psychology: General, 129*(2), 220.

Intraub, H. (2010). Rethinking scene perception: A multisource model. In B. H. Ross (Ed.), *Psychology of learning and motivation* (Vol. *52*, pp. 231–264). New York: Academic Press.

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

Intraub, H. (2020). Searching for boundary extension. *Current Biology, 30*(24), R1463–R1464.

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.

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*(2), 118–134.

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

Josephs, E. L., & Konkle, T. (2019). Perceptual dissociations among views of objects, scenes, and reachable spaces. *Journal of Experimental Psychology: Human Perception and Performance, 45*(6), 715.

Josephs, E. L., & Konkle, T. (2020). Large-scale dissociations between views of objects, scenes, and reachable-scale environments in visual cortex. *Proceedings of the National Academy of Sciences, USA, 117*(47), 29354–29362.

Knapen, T. (2021). Topographic connectivity reveals task-dependent retinotopic processing throughout the human brain. *Proceedings of the National Academy of Sciences, USA, 118*(2), e2017032118.

Konkle, T., & Oliva, A. (2007). Normative representation of objects: Evidence for an ecological bias in object perception and memory. *Proceedings of the Annual Meeting of the Cognitive Science Society, 29*, 407–412.

Konkle, T., & Oliva, A. (2011). Canonical visual size for real-world objects. *Journal of Experimental Psychology: Human Perception and Performance, 37*(1), 23.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. (2017). lmertest package: Tests in linear mixed effects models. *Journal of Statistical Software, 82*(1), 1–26.

Leyssen, M. H., Linsen, S., Sammartino, J., & Palmer, S. E. (2012). Aesthetic preference for spatial composition in multiobject pictures. *i-Perception, 3*(1), 25–49.

Lin, F., Hafri, A., & Bonner, M. F. (2022). Scene memories are biased toward high-probability views. *Journal of Experimental Psychology: Human Perception and Performance, 48*(10), 1116.

McDunn, B. A., Brown, J. M., Hale, R. G., & Siddiqui, A. P. (2016). Disentangling boundary extension and normalization of view memory for scenes. *Visual Cognition, 24*(5-6), 356–368.

Mullin, C. R., & Steeves, J. K. (2013). Consecutive TMS-fMRI reveals an inverse relationship in bold signal between object and scene processing. *Journal of Neuroscience, 33*(49), 19243–19249.

Palmer, E. M., Horowitz, T. S., Torralba, A., & Wolfe, J. M. (2011). What are the shapes of response time distributions in visual search? *Journal of Experimental Psychology: Human Perception and Performance, 37*(1), 58.

Palmer, S. (1981). Canonical perspective and the perception of objects. In A. D. Baddeley, & J. Long (Eds.), *Attention and performance IX* (pp. 135–151). Mahwah, NJ: Lawrence Erlbaum Associates.

Palmer, S. E., Schloss, K. B., & Sammartino, J. (2013). Visual aesthetics and human preference. *Annual Review of Psychology, 64*, 77–107.

Park, J., Josephs, E., & Konkle, T. (2022). Ramp-shaped neural tuning supports graded population-level representation of the object-to-scene continuum. *Scientific Reports, 12*(1), 18081.

Park, J., & Park, S. (2020). Coding of navigational distance and functional constraint of boundaries in the human scene-selective cortex. *Journal of Neuroscience, 40*(18), 3621–3630.

Park, J., Soucy, E., Segawa, J., Mair, R., & Konkle, T. (2023). Ultra-wide angle neuroimaging: Insights into immersive scene representation. *bioRxiv,* http://doi.org/10.1101/2023.05.14.540275.

R Core Team. (2016). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing (https://www.R-project.org/).

Ratcliff, R. (1979). Group reaction time distributions and an analysis of distribution statistics. *Psychological Bulletin, 86*(3), 446.

Strong, R. W., & Alvarez, G. (2019). Using simulation and resampling to improve the statistical power and reproducibility of psychological research. *PsyArXiv,* https://doi.org/10.31234/osf.io/2bt6q.