AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Experimental Psychology:
Human Perception and Performance

ISSN: 0096-1523

2022, Vol. 48, No. 10, 1116–1129
https://doi.org/10.1037/xhp0001045

Scene Memories Are Biased Toward High-Probability Views

Feikai Lin, Alon Hafri, and Michael F. Bonner
Department of Cognitive Science, Johns Hopkins University

Visual scenes are often remembered as if they were observed from a different viewpoint. Some scenes are remembered as farther than they appeared, and others as closer. These memory distortions—also known as boundary extension and contraction—are strikingly consistent for a given scene, but their cause remains unknown. We tested whether these distortions can be explained by an inferential process that adjusts scene memories toward high-probability views, using viewing depth as a test case. We first carried out a large-scale analysis of depth maps of natural indoor scenes to quantify the statistical probability of views in depth. We then assessed human observers' memory for these scenes at various depths and found that viewpoint judgments were consistently biased toward the modal depth, even when just a few seconds elapsed between viewing and reporting. Thus, scenes closer than the modal depth showed a boundary-extension bias (remembered as farther-away), and scenes farther than the modal depth showed a boundary-contraction bias (remembered as closer). By contrast, scenes at the modal depth did not elicit a consistent bias in either direction. This same pattern of results was observed in a follow-up experiment using tightly controlled stimuli from virtual environments. Together, these findings show that scene memories are biased toward statistically probable views, which may serve to increase the accuracy of noisy or incomplete scene representations.

Public Significance Statement

Humans make striking errors when recalling how they viewed of a visual scene, with some scenes misremembered as farther than they appeared and others as closer (called boundary extension and contraction). For example, one might recall an entire street scene after briefly seeing a close view of a fire hydrant (i.e., boundary extension), or one might recall a close view of a building after briefly seeing it from a distance (i.e., boundary contraction). These memory errors are highly consistent for specific scenes, yet their explanation is still unknown. By combining a novel analysis of depth information in visual scenes with behavioral experiments on scene memory, we found evidence that observers' memories for scenes are biased toward the usual depth of those scenes. Our findings suggest that these memory errors arise from an inferential process that biases memories to high-probability views, which may serve to increase the accuracy of memory in noisy conditions.

Keywords: boundary extension and contraction, natural statistics, scene memory, scene perception

Memories for visual scenes are prone to systematic errors. One striking form of memory error can be detected in how observers recall their viewpoint in a scene, with some scenes misremembered at farther viewpoints (an effect called boundary extension) and others misremembered at closer viewpoints (an effect called boundary contraction; Bainbridge & Baker, 2020a; Intraub & Richardson, 1989). The first reports of these memory biases detected far more boundary extension than contraction and proposed that the underlying mechanism was an anticipatory scene-construction process that extrapolates beyond the boundaries of the immediate view (Intraub et al., 1992; Intraub & Richardson, 1989; see for review: Intraub, 2010). However, recent studies with larger and more diverse stimulus sets have revealed that scene boundaries are equally likely to extend or contract in memory in a manner that is consistent across observers, demonstrating that these effects are inherently bidirectional and cannot be fully explained by an anticipatory scene-construction process (Bainbridge & Baker, 2020a; Park et al., 2021; cf. Bainbridge & Baker, 2020b; Intraub, 2020; for discussion). Thus, boundary extension and contraction (or together, boundary transformation) appear to reflect a central aspect of scene processing in the human mind whose underlying cause remains unknown.

This article was published Online First August 18, 2022.

Michael F. Bonner https://orcid.org/0000-0002-4992-674X

Feikai Lin is now at the Institute of Neuroscience, CAS Center for Excellence in Brain Science and Intelligence Technology, Chinese Academy of Sciences. Alon Hafri is now at the Department of Linguistics and Cognitive Science, University of Delaware.

We thank Eric Elmoznino for help with data collection. This work was partially supported by National Science Foundation Social, Behavioral and Economic Sciences Postdoctoral Research Fellowship SMA-2105228 awarded to Alon Hafri.

Correspondence concerning this article should be addressed to Michael F. Bonner, Department of Cognitive Science, Johns Hopkins University, 3400 North Charles Street, Baltimore, MD 21218, United States. Email: mfbonner@jhu.edu

Theories in other areas of memory and perception suggest a candidate mechanism for boundary extension and contraction: Specifically, these biases could arise from a prior-based inference mechanism that fuses scene representations with prior knowledge of the visual environment (Griffiths et al., 2008; Kersten et al., 2004; Knill & Richards, 1996; Petzschner et al., 2015). In many scenarios, these inferences will tend to *increase* representational accuracy, because they leverage prior knowledge of the statistical regularities of the environment to push noisy stimulus representations toward high-probability states. For example, such a process could facilitate robust object recognition: On a rainy evening with low visibility, an observer may still infer that an oblong shape on the road is a car. However, these prior-based inferences can also produce systematic patterns of *errors*—in particular, errors in the direction of high-probability representational states. Might this inference process explain the viewpoint biases of boundary extension and contraction, and if so, what scene priors underlie these inferences?

One candidate prior is viewing depth: In images of real-world scenes, views at some depths are more likely than others. For example, images of indoor environments regularly capture the full layout of the room rather than a close-up view of a wall. Such regularities constitute a probability distribution over views. While previous work points to the importance of viewing depth as a mediating factor for boundary extension and contraction (Bainbridge & Baker, 2020a; Bertamini et al., 2005; Hafri et al., in press; Park et al., 2021), no studies have directly tested whether these biases result from memories being drawn toward views at high-probability depths. Furthermore, although it has previously been speculated that the statistical regularities of view depths influence viewpoint judgments (Konkle & Oliva, 2007), no studies have systematically quantified these statistical regularities. Thus, it remains unknown whether boundary transformation is related to depth probability. Alternatively, it is possible that boundary transformation and depth probability have no underlying relationship or that their relationship is inconsistent with our prior-based hypothesis. For example, one alternative possibility is that high-probability views are more familiar, such that observers are better at extrapolating the scene content beyond the boundaries of these views. This would predict that the strongest boundary extension effects would be observed for high-probability views (Hubbard et al., 2010; Intraub, 2002), which would be inconsistent with our hypothesis.

To test our prior-based hypothesis, we developed a novel analysis of a large-scale computer-vision dataset of indoor scenes to quantify the natural statistics of scene views. We then performed three behavioral experiments to assess memory biases: Experiments 1a and 1b with natural scenes and Experiment 2 with virtual scenes. To anticipate, our findings show that viewpoint judgments of scenes are consistently biased toward statistically probable views in depth.

Experiment 1a: Natural Scenes

We first needed an approach for quantifying the statistical regularities of scene views. For our measure of interest, we examined the mean depth across all pixels of an image, which we call the image depth. We reasoned that image depth would be a useful summary statistic of the depth regularities that might underlie viewpoint priors.

Method

Quantifying the Probability of Image Depths

We took advantage of the publicly available SUN RGB-D dataset to quantify the statistical regularities of depth for images of indoor scenes (Song et al., 2015). This dataset contains 10,335 images from 47 scene categories, and each image has a corresponding 3D depth map with a real-world depth value at each pixel (Figure 1A). The images were collected by experimenters as they navigated around real-world environments while wearing a camera and a depth-sensing instrument (e.g., Kinect v2). The environments were mostly universities, houses, and furniture stores, and the images were labeled with the scene categories within these environments (e.g., lecture hall, kitchen, bathroom). Although there were no explicit constraints on the sampling of these images, the sampling procedure, nonetheless, reflected the inherent regularities of scene views collected during navigation in indoor environments. We focused on the top 15 scene categories with the largest numbers of images, and we examined the refined depth maps for these images in the SUN RGB-D dataset. For each image in each selected category, we obtained a single summary statistic of image depth by calculating the mean depth across all pixels. We then used kernel density estimation to fit a probability distribution over the depths of all images in each scene category, and we also fit a global probability distribution over all images in the SUN RGB-D dataset (Figures 1A). Our motivation for examining depth distributions for images from different categories was to ensure that our results generalized across categories and were not driven by any single scene category. Our experiments were not designed to test category-specific effects, and, indeed, most categories showed highly overlapping depth distributions, apart from two categories, Bathroom and Furniture Store. We later examined these two categories to explore whether the nature of the depth effect on boundary transformation is general or category-specific (see Results section of Experiment 1b).

Stimuli

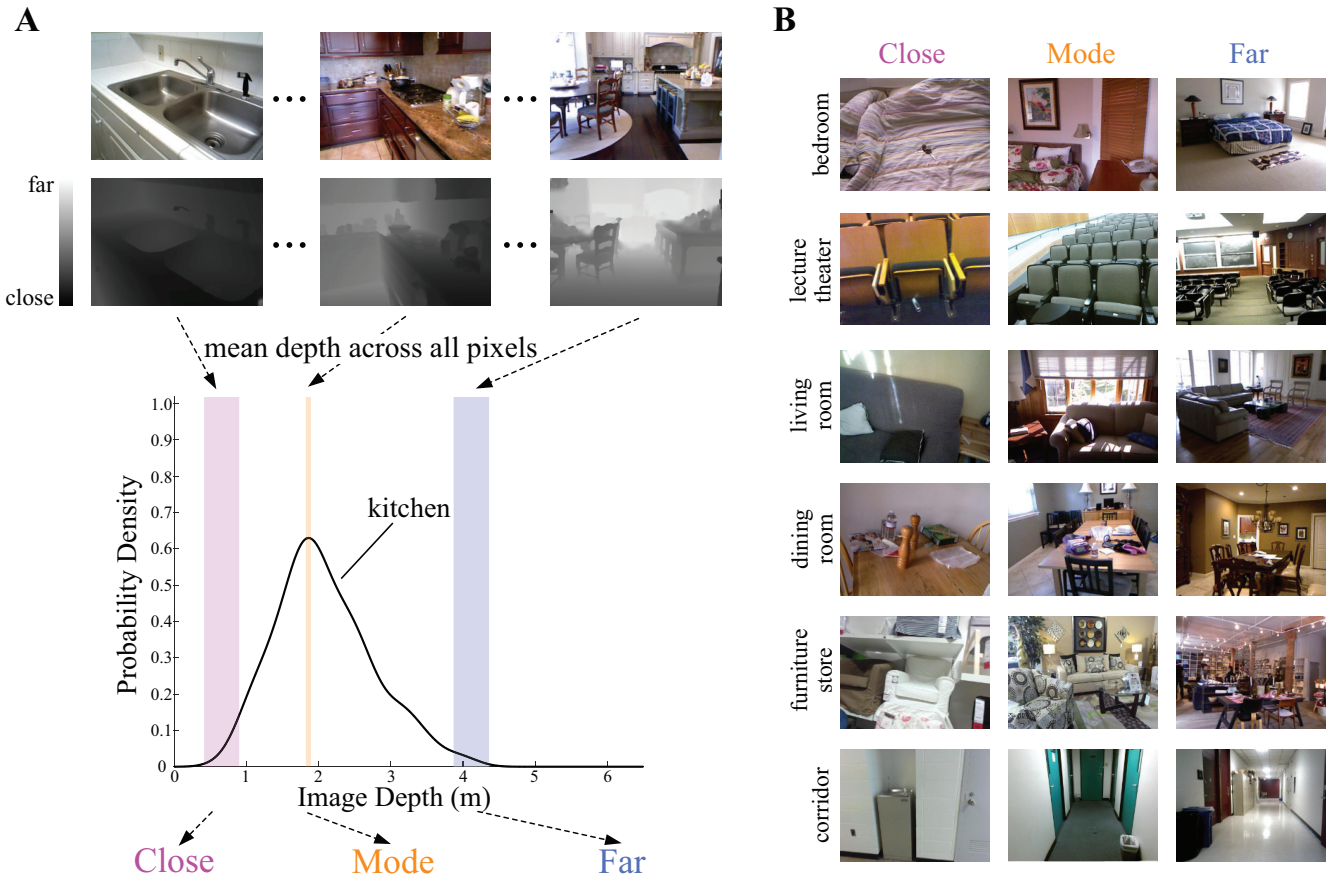
We sorted image stimuli into three experimental conditions, Close, Mode, and Far, based on where they fell along the probability distribution of image depth of their scene categories (see Figure 1). To maximize the diversity of our stimuli, we ensured that no condition contained multiple images of the same exact scene at different views. Close images were near the minimum image depth, Mode images were near the statistical mode, and Far images were near the maximum. We selected three images in each of three conditions (Close, Mode, and Far) from 13 scene categories, yielding a total of 117 images. These images were presented at 640 × 480 pixels in our online rapid serial visual presentation (RSVP) experiments.

Participants

Participants were recruited on the online experiment platform Prolific in 2020 with the following criteria: residency in the

Figure 1

Selection of Natural-Scene Stimuli Based on Probability Distributions of Image Depth



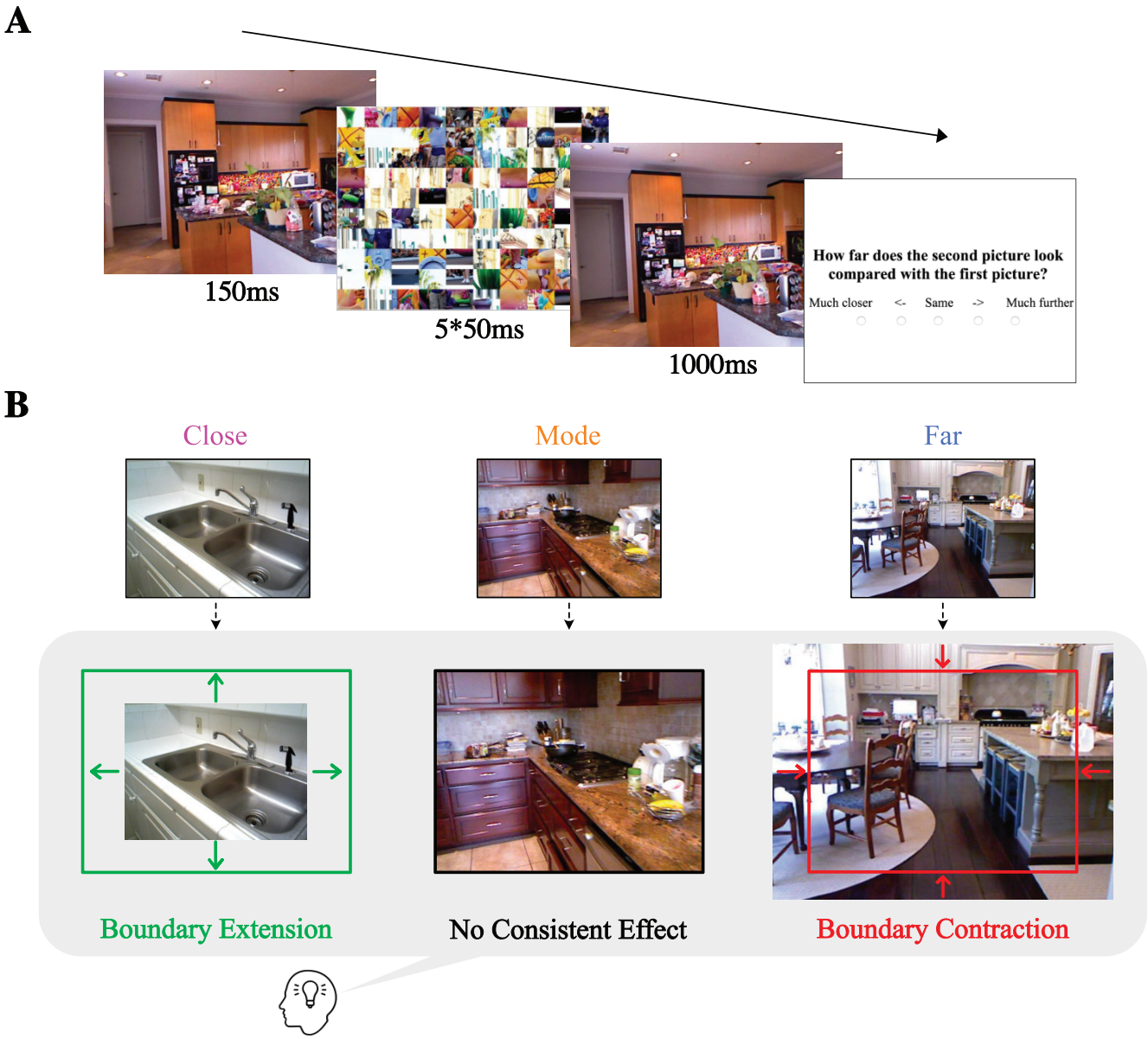
Note. (A) The statistical probabilities of image depth were quantified for images of indoor scenes in the SUN RGB-D dataset, which contains ground-truth depth maps for all images (Song et al., 2015). For each image, a summary statistic of image depth was computed by averaging across all pixels in its depth map. For each scene category, image depth values for all images in that category were used to generate a probability distribution. This panel shows an example for the category kitchen. Stimuli were selected for three experimental conditions (Close, Mode, and Far) based on where they fell along the probability distribution of image depth. Shaded areas indicate the ranges of image depths for stimuli in the Close, Mode, and Far conditions. This process was repeated for all scene categories. (B) Example stimuli for experimental conditions (Close, Mode, and Far) are shown for six scene categories. Example stimuli from all categories are shown in Appendix Figure A1. See the online article for the color version of this figure.

United States, normal or corrected-to-normal vision, total completed submissions larger than 50, and approval rate higher than 95%. The target population for all experiments in this article was sighted adults. The recruitment criteria were the same for all experiments. These criteria were designed to select for motivated online study participants with normal or corrected-to-normal vision, and beyond these criteria, we have no reason to believe that the results reported here depend on other characteristics of the participants. We collected 39 participants in this pilot study, and we used the data to conduct power analyses and determine sample sizes for follow-up experiments (see Method section in Experiment 1b for more details). Our experiments included simple attention-check trials (described below), and participants were excluded from further analyses if they made at least two errors on these trials. We additionally removed participants before further analyses if they gave the same response on more than 90% of experimental trials. Overall, data from 27 participants in Experiment 1a (14 males and 13 females; age range: 20–60 years, $M = 33.70$, $SD = 11.04$) were taken into hypothesis testing. All studies were approved by the Johns Hopkins University Institutional Review Board.

RSVP Experiment

We assessed viewpoint biases using an RSVP paradigm (Figure 2A), which has previously been shown to elicit consistent effects of boundary extension and contraction (Bainbridge & Baker, 2020a; Hafri et al., in press; Intraub & Dickinson, 2008; Park et al., 2021). For the RSVP experiment, stimuli were randomly distributed into nine blocks, with each scene category only appearing once in each block. The order of trials within blocks was randomized. The RSVP paradigm was adapted from Bainbridge & Baker, 2020a. In experimental trials (Figure 2A), participants saw a fixation cross for 300 ms, then an image for 150 ms, followed by 250 ms of a dynamic mask (five mosaic-scrambled images not containing the preceding

Figure 2
Behavioral Task for Assessing Boundary Transformation



Note. (A) A rapid serial visual presentation (RSVP) paradigm was used to quantify boundary-transformation effects for our stimuli. In each trial, participants saw a briefly presented image, followed by a dynamic mask, and then a second presentation of the exact same image. They were then asked to judge, using a 5-point scale, how far the second image looked compared with the first image. Crucially, participants were not told that the second image would always be identical to the first image. (B) Our hypothesis was that boundary-transformation effects would reflect biases toward statistically probable views in depth, which would lead to boundary extension for the Close scenes, boundary contraction for the Far scenes, and no consistent bias in either direction for the Mode scenes. See the online article for the color version of this figure.

experimental image, presented for 50 ms each), and then the original image again (although, importantly, participants were never told that the images might be the same). After 1,000 ms, the image disappeared and a response screen appeared asking how far the second picture looked compared with the first one, along with five options (*much closer*, *a bit closer*, *the same*, *a bit further*, and *much further*). On average, participants responded within three seconds. After they responded, the next trial started after a 500-ms delay.

Each block of the experiment also included one attention-check trial to ensure that participants paid attention to both the first and the second images. In attention-check trials, the main procedure was the same, except that the second image was either identical to the first image or a completely different scene, and on the response screen, participants were asked whether the second image was the same as the first one. The question prompt on attention-check trials was also highlighted in a bold color to emphasize that a different question was

being asked. Before the actual experiments started, participants were given detailed instructions and performed practice trials with feedback. During the actual experiments, participants were told their performance on the attention-check trials at the end of each block.

Boundary Transformation Scoring

In experimental trials, participants responded on a five-option scale. We converted responses to numerical values that we called boundary-transformation scores: *much closer* (scored as −2), *a bit closer* (scored as −1), *the same* (scored as 0), *a bit further* (scored as +1), and *much further* (scored as +2). Negative scores indicate boundary extension, and positive scores indicate boundary contraction. The final boundary transformation score for each image was the average of participants' responses for that image.

Analyses of Between-Experiment and Within-Experiment Consistency

All statistical analyses were conducted in RStudio 1.3.959. We examined the between-experiment consistency of boundary-transformation effects for natural scenes in Experiments 1a and 1b. Because the image instances in Experiments 1a and 1b were different, we examined consistency for the 13 scene categories that appeared in both experiments (see Method section in Experiment 1b). We computed a mean boundary-transformation score across stimuli for each scene category in each depth condition, and we computed the Spearman's rank correlation ρ of these boundary-transformation scores across experiments. We also evaluated the within-experiment consistency of boundary-transformation scores in all experiments with a split-half procedure. For each experiment, across 1,000 iterations, we randomly split the participants into two groups. We computed the boundary-transformation score for each image within each group and then computed the Spearman's rank correlation of these image-wise boundary-transformation scores across the two groups. The Spearman's rank correlation was then corrected based on the Spearman-Brown prediction formula (Spearman's rank ρ'), as in Bainbridge & Baker, 2020a.

Analyses of Boundary Transformation Scores

To evaluate the effects of the depth conditions on the boundary-transformation scores, we performed one-way analyses of variance (ANOVAs) separately on the item-level (i.e., image-level) and participant-level boundary-transformation scores. We applied Greenhouse-Geisser correction to the degrees of freedom in ANOVAs when the assumption of sphericity was violated in Mauchly's test ($p < .05$). We followed up the ANOVAs with post hoc *t* tests on each depth condition with Bonferroni correction. To complement these analyses, we also calculated Bayes factor *t* tests using the R package BayesFactor with the function ttestBF and the default medium prior of sqrt(2)/2. Bayes Factors were not corrected for multiple comparisons. We looked for evidence in favor of the alternative hypothesis (i.e., BF_{10}). Values greater than 1 indicate evidence for the alternative, and values less than 1 indicate evidence in favor of the null hypothesis of no difference.

Transparency and Openness

All image stimuli (selected SUN RGB-D images and generated virtual scenes), data (participants' responses in the behavioral experiments), and analysis code are available in a repository on Open Science Foundation (https://osf.io/y9pue). None of the studies was preregistered.

Results

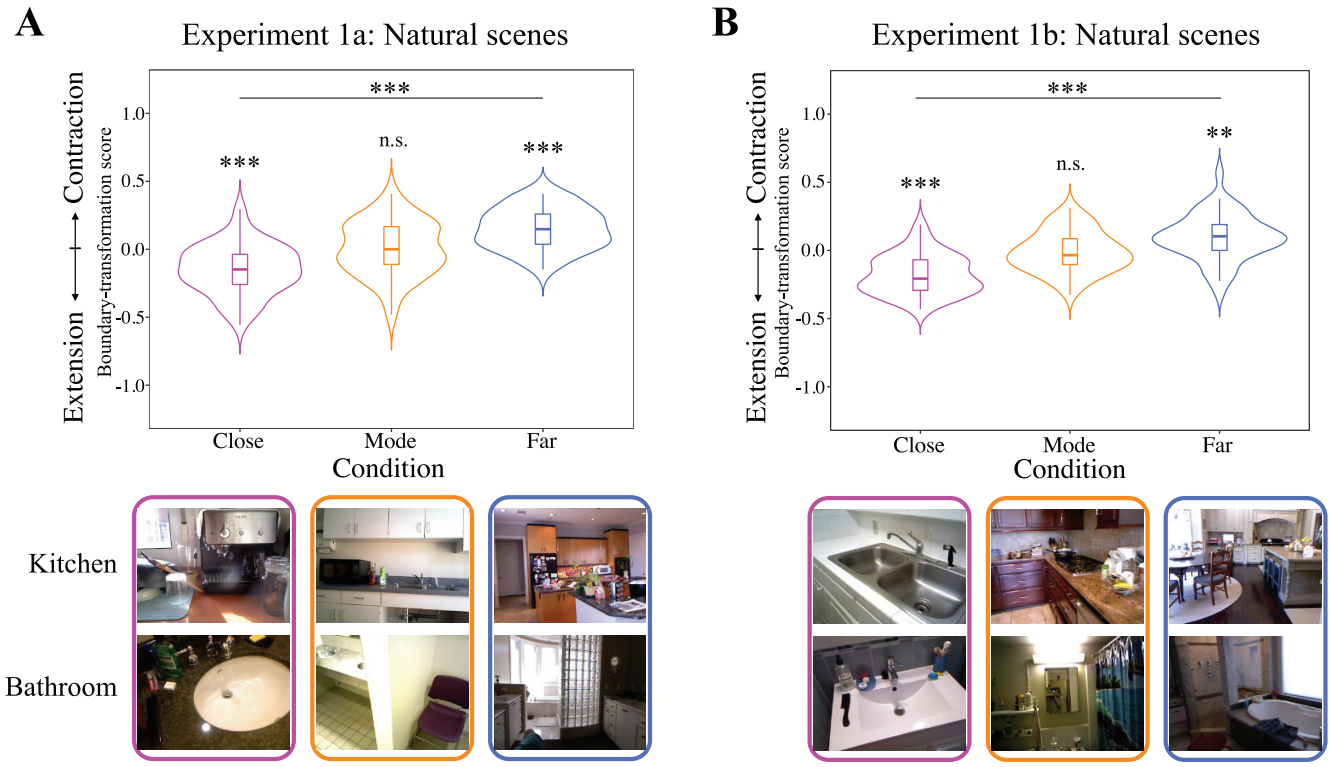
We first analyzed the reliability of boundary-transformation scores within Experiment 1a (Appendix Figure A2B). The analysis showed that boundary-transformation scores were consistent across groups of participants within the experiment (ρ' = .66, $p <$ .001).

For the main analysis of interest, we compared scores across the three depth conditions to test our central prediction that viewpoint judgments would be biased toward the modal depth, thus resulting in a boundary-extension bias for the Close condition, a boundary-contraction bias for the Far condition, and no consistent bias for the Mode condition (Figure 2B). We report analyses examining variance for both image items and participants (Figure 3A). The item-level ANOVA showed a significant effect of depth condition, $F(2, 114) = 26.86$, $p < .001$, $\eta^2 = .32$. Post hoc *t* tests with Bonferroni correction showed that boundary transformation scores for images in both the Close condition ($M = -.15$) and Far condition ($M = .15$) differed significantly from zero and that scores in the Mode condition ($M = .01$) did not differ from zero, as indicated by the Bayes Factor close to zero for the Mode condition (Close: $t[38] = -4.97$, $p < .001$, $d = .80$, $BF_{10} = 1.43e3$; Far: $t[38] = 6.25$, $p < .001$, $d = 1.00$, $BF_{10} = 6.03e4$; Mode: $t[38] = .18$, $p = .999$, $d = .03$, $BF_{10} = .18$). The participant-level ANOVA with Greenhouse-Geisser correction also showed a significant effect of depth condition, $F(1.22, 31.79) = 14.17$, $p < .001$, $\eta_p^2 = .35$. Corrected post hoc *t* tests and Bayes Factor revealed that boundary-transformation scores for images in the Close condition ($M = -.15$) and Far condition ($M = .15$), but not the Mode condition ($M = .01$), differed significantly from zero (Close: $t[26] = -4.74$, $p < .001$, $d = .91$, $BF_{10} = 3.83e2$; Far: $t[26] = 2.67$, $p < .05$, $d = .51$, $BF_{10} = 3.74$; Mode: $t[26] = .21$, $p = .999$, $d = .04$, $BF_{10} = .21$). In sum, these results indicated that viewpoint judgments for memories of natural scenes are biased toward high-probability views in depth.

Although we randomized the order of stimuli across and within blocks, it is possible that participants could have changed their pattern of responses over the course of the experiments as they saw stimuli from various depth conditions, which would suggest that their responses were not driven by existing priors but rather by ad hoc priors specific to our experimental stimuli. If that were the case, we would expect the boundary transformation scores for the Close and Far conditions to increase over the course of the experiment. To test this possibility, we examined participants' responses in the first and second halves of the experiment. To do this, we split the nine experimental blocks into two halves (i.e., the first five blocks and the last four blocks), and we calculated each participant's average response for the stimuli in each depth condition for each half, as in our initial analyses using all blocks. We then performed a two-way ANOVA with Greenhouse-Geisser correction on the participant-level boundary-transformation scores with depth condition (Close, Mode, and Far) and time (first half and second half) as categorical variables. We found a significant main effect of depth condition, $F(1.25, 32.43) = 13.19$, $p < .001$, $\eta_p^2 = .34$.

Figure 3

Boundary-Transformation Effects Revealed Biases Toward Statistically Probable Image Depths



Note. (A) Violin plots of item-level boundary-transformation scores for the pilot study, Experiment 1a, showed that memory judgments were biased toward the modal depth. Specifically, stimuli in the Close condition mostly elicited boundary extension, as they were reported as being closer than recalled. In contrast, stimuli in the Far condition elicited boundary contraction, as they were reported as being farther than recalled. Finally, stimuli in the Mode condition elicited no consistent bias in either direction. These violin plots show the full distribution of boundary transformation scores for all stimuli, and they are overlayed with a box-and-whisker plot showing the median (midline), the first and third quartiles (upper and lower hinges), and 1.5 times the interquartile range (whiskers). Two example stimuli for each depth condition are shown below the violin plot. (B) Violin plots of item-level boundary transformation scores for Experiment 1b. The findings were consistent across both experiments. Plotting conventions for the violin plot are the same as in panel A. Example images are shown at the bottom of the figure. Asterisks above each line indicate significance for ANOVA, and those above each violin indicate significance for a two-tailed t test against zero. n.s. = not significant ($p > .05$). See the online article for the color version of this figure.

** $p < .01$. *** $p < .001$.

Although we did not see a main effect of time, $F(1, 26) = 1.27$, $p = .270$, $\eta_p^2 = .05$, we did find a significant interaction of depth condition and time, $F(2, 52) = 4.13$, $p = .022$, $\eta_p^2 = .14$. However, the nature of this interaction was inconsistent with the prediction that boundary transformation scores would increase over the course of the experiment. Rather, the boundary transformation scores appeared to get weaker over the experiment, particularly for the Close condition (Close condition first half $M = -.21$ and second half $M = -.08$). In sum, this analysis suggests that the boundary transformation effects observed here did not result from participants learning the statistics of the experimental stimulus set over the course of the experiment.

Experiment 1b: Natural Scenes

Experiment 1b was a conceptual replication of Experiment 1a with new stimuli and new participants.

Method

Stimuli

Based on the probability distributions over the depths for all scene categories in the SUN RGB-D dataset, we followed the same procedure described for Experiment 1a and sorted new sets of natural images into three experimental conditions, Close, Mode, and Far. We selected three images in each of three conditions (Close, Mode, and Far) from 15 scene categories, yielding 45 stimuli per experimental condition and a total of 135 images. None of the stimuli overlapped with the stimuli used in Experiment 1a. Example stimuli from each category are shown in Appendix Figure A1.

Sample-Size Determination and Power Analysis

Sample sizes for this experiment and for Experiment 2 were determined a priori based on a power analysis of data from Experiment 1a. Specifically, we conducted a power analysis on the

weakest of the Close and Far effects in the post hoc one-sample *t* tests in Experiment 1a, because these were the conditions that were expected to differ from zero. Using G*Power 3.1 software, we input the effect size for the Experiment 1a Far condition (i.e., the smaller of the Close and Far effects, $d = .51$) and the α error level ($\alpha = .05$). This analysis showed that we would need at least 42 participants to have 90% power to detect an effect size equal to or greater than that observed in Experiment 1a. Given the relatively high exclusion rate of 31% for Experiment 1a, we collected data from 100 participants each in Experiments 1b and 2 to ensure that we would have sufficient power to detect the effects of interest after exclusions.

Participants

A new group of 100 participants was recruited on Prolific in 2020 and 2022 using the same criteria as in Experiments 1a. Exclusion criteria and experimental procedures were the same as Experiment 1a. As a result, data from 58 participants (31 males and 27 females; age range: 19–74 years, $M = 40.16$, $SD = 13.83$) were taken into further analyses.

Procedure and Analyses

The RSVP experiment and data analysis procedures were the same as in Experiment 1a.

Analyses of Categorical Effects With Bathroom and Furniture Store

To explore the categorical effects on boundary transformation scores, we focused on images in different scene categories but with similar depths. Comparing the image depth distributions for all categories and the global depth distribution (Figure 4A), we found that the distributions of bathroom and furniture store diverged the most from others. Combining images from Experiments 1a and 1b, we selected bathroom images from the Mode condition and then found images from other scene categories whose depths were within .06 m of the bathroom images. This process yielded six images in the Bathroom Mode condition ($M = 1.60$, $SD = .001$) and six images in the Others condition (four images of dining area in Close condition, one image of furniture store in Close condition, and one image of lecture theater in Close condition, $M = 1.57$, $SD = .027$). We repeated this process for furniture store and obtained six images in the group of Furniture-Store Mode ($M = 2.65$, $SD = .003$) and another six images in Others (five images of dining room in Mode condition and one image of bathroom in Far condition, $M = 2.63$, $SD = .032$). We used two-sample *t* tests to compare the boundary transformation scores in each pair of groups. Note that degrees of freedom were approximated using the Welch-Satterthwaite equation, assuming two independent samples with unequal variance.

Results

We first analyzed the reliability of boundary transformation scores across Experiments 1a and 1b and within Experiment 1b (Appendix Figure A2A and A2B; see Method section in Experiment 1a). The analysis showed that boundary transformation scores were consistent across the two experiments ($\rho = .72$, $p < .001$) and across groups of participants within Experiment 1b ($\rho' = .80$, $p < .001$).

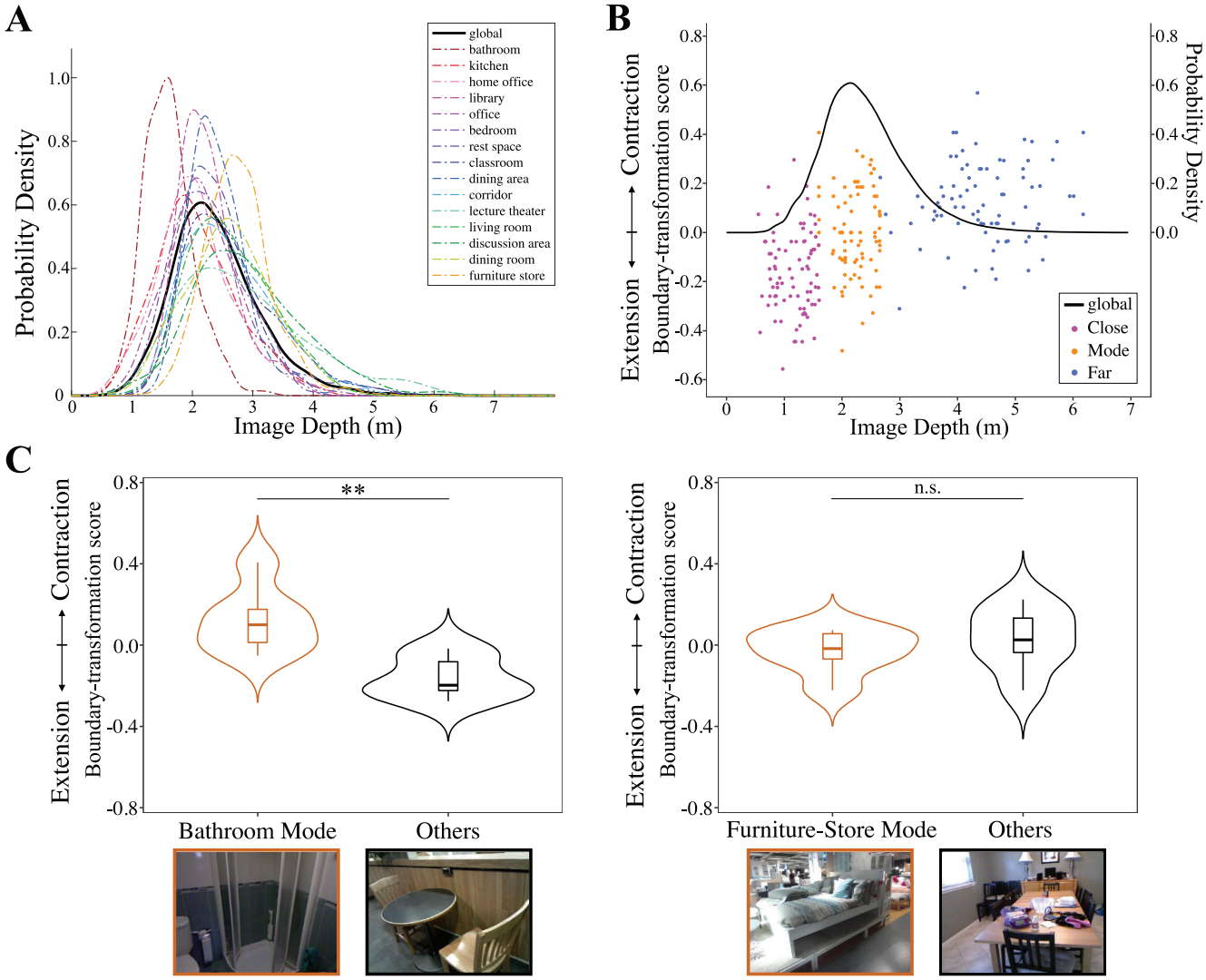
For the main analysis of interest, we again compared boundary transformation scores across the three depth conditions to test our central prediction. We report analyses examining variance for both image items and participants (Figure 3B). The item-level ANOVA confirmed a significant effect of depth condition, $F(2, 132) = 35.47$, $p < .001$, $\eta^2 = .35$. Post hoc *t* tests with Bonferroni correction showed that boundary transformation scores in both the Close condition ($M = -.18$) and Far condition ($M = .10$) differed significantly from zero and that scores in the Mode condition ($M = -.01$) did not differ from zero, as indicated by the Bayes Factor close to zero for the Mode condition (Close: $t[44] = -8.16$, $p < .001$, $d = 1.22$, $BF_{10} = 4.47e7$; Far: $t[44] = 3.82$, $p < .01$, $d = .57$, $BF_{10} = 6.49e1$; Mode: $t[44] = -.50$, $p = .999$, $d = .07$, $BF_{10} = .18$). Additionally, the participant-level ANOVA confirmed a significant effect of depth condition, $F(1.47, 83.97) = 46.23$, $p < .001$, $\eta_p^2 = .45$. Corrected post hoc *t* tests and Bayes Factor showed that boundary transformation scores for images in the Close condition ($M = -.18$) and Far condition ($M = .10$) differed significantly from zero, while the Mode condition ($M = -.01$) did not differ from zero (Close: $t[57] = -5.33$, $p < .001$, $d = .70$, $BF_{10} = 9.49e3$; Far: $t[57] = 2.66$, $p < .05$, $d = .35$, $BF_{10} = 3.54$; Mode: $t[57] = -.37$, $p = .999$, $d = .05$, $BF_{10} = .15$). Overall, our results consistently indicate that viewpoint judgments for memories of natural scenes are biased toward high-probability views in depth.

We also wanted to test whether participants changed their pattern of responses over the course of Experiment 1b. As in Experiment 1a, we split the nine experimental blocks into two halves (i.e., the first five blocks and the last four blocks), and we performed a two-way ANOVA with Greenhouse-Geisser correction on the participant-level boundary transformation scores with depth condition (Close, Mode, and Far) and time (first half and second half) as categorical variables. We found a significant main effect of depth condition, $F(1.48, 84.21) = 45.75$, $p < .001$, $\eta_p^2 = .45$, no effect of time, $F(1, 57) = 2.71$, $p = .105$, $\eta_p^2 = .05$, and no interaction of depth and time, $F(2, 114) = .02$, $p = .978$, $\eta_p^2 = .00$. Thus, participants' responses were stable over the course of the experiment, and our main results were unlikely to be explained by ad hoc priors that participants learned during the experiment.

Note that our findings here and in Experiment 1a cannot resolve whether boundary transformation is driven by a single prior for all indoor scenes or by distinct category-specific priors (e.g., one might expect that the relationship between boundary transformation and depth differs across scene categories with seemingly different depth distributions, like bathrooms and furniture stores). Although we selected our stimuli using separate depth distributions for different categories in the dataset, these distributions are highly overlapping (Figure 4A). Indeed, the boundary transformation scores for our stimuli were broadly consistent with a bias toward the global mode of image depth across all indoor scenes in SUN RGB-D. Specifically, on average there was no bias near the peak of the global probability distribution, but there was boundary extension for stimuli near the lower tail and boundary contraction for stimuli near the upper tail (Figure 4B). Nevertheless, two of our categories—bathroom and furniture store—appeared to diverge from the others in their category-specific depth distributions (Figure 4A), which presents an opportunity to look for evidence of category-specific priors.

To explore this possibility, we asked whether images from the bathroom and furniture-store Modes would show different boundary-transformation effects compared with images from other scene categories near the same depths. If so, it would be suggestive of category-specific priors. With combined data from Experiments 1a

Figure 4
Global and Category-Specific Probability Distributions of Image Depth



Note. (A) Probability distributions of depth were highly overlapping across different categories of indoor scenes (colored dash lines), and most categories did not diverge from the global probability distribution computed across all images in SUN RGB-D (solid black line). (B) Item-wise boundary-transformation scores for all images in the natural-scene experiments are plotted along with the global distribution of image depth. Each point represents the average score for a single image. About 29% of variance in boundary transformation scores could be attributed to image depth ($R^2 = 0.29$). Note that the boundary transformation effects appeared to be broadly consistent with a bias toward the global mode of image depth. (C) Violin plots show item-level boundary transformation scores for Bathroom Mode (left) and Furniture-Store Mode (right) against depth-matched images in the corresponding Others condition (see Method section above). One example stimulus in each group is shown below the violins. Plotting conventions for the violin plots are the same as in Figure 3. Asterisks above the line indicate significance for two-tailed, two-sample *t*-tests. n.s. = not significant ($p > .05$). See the online article for the color version of this figure.
** $p < .01$.

and 1b, we separately compared boundary transformation scores for the bathroom and furniture-store Mode conditions with all other scenes whose image depths were within .06 m of the mean depths for these conditions (see the Method section). Two-sample *t* tests showed that when compared with the bathroom Mode condition ($M = .12$), images from other scene categories had lower boundary-transformation scores ($M = -.16$; i.e., more boundary extension), $t(8.41) = 3.57$, $p < .01$, $d = 2.06$ (Figure 4C, left). However, boundary-transformation scores for the furniture-store Mode condition ($M = -.03$) were not significantly different from boundary transformation scores for other categories at a similar depth ($M = .03$), $t(8.95) = -.74$, $p = .478$, $d = .43$ (Figure 4C, right). One reason that we might observe this effect for the bathroom category but not the furniture store category is that in the comparison for the bathroom Mode, the nearby images from other scene categories were all in the Close condition, but in the

comparison for the furniture-store Mode, nearby images from other scene categories were mostly from the Mode condition (apart from one image in the Far condition). Nonetheless, the finding for the bathroom condition provides some suggestive evidence that when category-specific priors strongly diverge from the global prior, it may be possible to detect category-specific patterns of boundary transformation effects. However, because only two of our categories showed any divergence from the global depth distribution, we cannot adjudicate between category-specific or more general priors. We thus leave this question for future work to address.

Experiment 2: Virtual Scenes

Although it was important to establish that this pattern of viewpoint biases could be observed with natural scene stimuli, a caveat of this approach was that our experimental conditions varied not only in depth but also in the exact environments in which the photographs were taken. Thus, although all three depth conditions contained images from each scene category (e.g., kitchen), each depth condition contained different instances of these categories (e.g., different kitchens appeared in the Close, Mode, and Far conditions). This left open the possibility that the viewpoint biases we observed were affected by other environment-specific factors. We addressed this concern by creating a set of realistic virtual environments and rendering images of these environments at various viewing distances. This allowed us to examine viewpoint biases at different distances while keeping the exact environment constant.

Method

Stimuli

Computer-generated virtual environments were constructed in Blender (Version 2.83.4). We created virtual environments for nine different scene categories that overlapped with the categories used in the natural scene experiments (e.g., living room, corridor, classroom). For some of these categories, we created multiple environments, yielding a total of 16 unique virtual environments. Example images of these environments can be seen in Appendix Figure A3. Images of virtual scenes were captured at three locations along a linear trajectory in each environment, with fixed virtual camera focal length and viewing angle. The camera locations were selected to capture the views of natural images from the same scene categories as in Experiments 1a and 1b. For example, a view of a virtual kitchen in the *Close* condition was designed to resemble a view of a real-world kitchen in the *Close* condition of the natural scene experiments. In total, 48 images were generated from 16 environments (one image for each of the three depth conditions). These images were originally captured in 1,600 × 1,200 pixels from Blender, and later presented at 640 × 480 pixels in the online RSVP experiment.

Participants

Sample size was determined a priori based on a power analysis of data from Experiment 1a (see Experiment 1b Method above). A new group of 100 participants was recruited on Prolific in 2020 and 2022 using the same criteria as in Experiments 1a and 1b. Exclusion criteria and experimental procedures were the same as in the previous experiments. As a result, data from 85 participants (52 males and 33 females; age range: 20–71 years, $M = 41.42$, $SD = 16.04$) were taken into further analyses.

RSVP Experiment

All 48 stimuli were randomly distributed into three blocks, with each environment appearing only once in each block. Each block contained 16 experimental trials, one attention-check trial, and four filler trials containing novel virtual environments that were different from the main stimulus set, which were included to reduce participants' expectations of the environments that they might encounter in each block. Otherwise, the experimental procedure was the same as in the natural scene experiments.

Analyses

Boundary-transformation scores were calculated as previously described. The analysis of within-experiment consistency and analyses to evaluate the effects of depth conditions were the same as in Experiments 1a and 1b.

Results

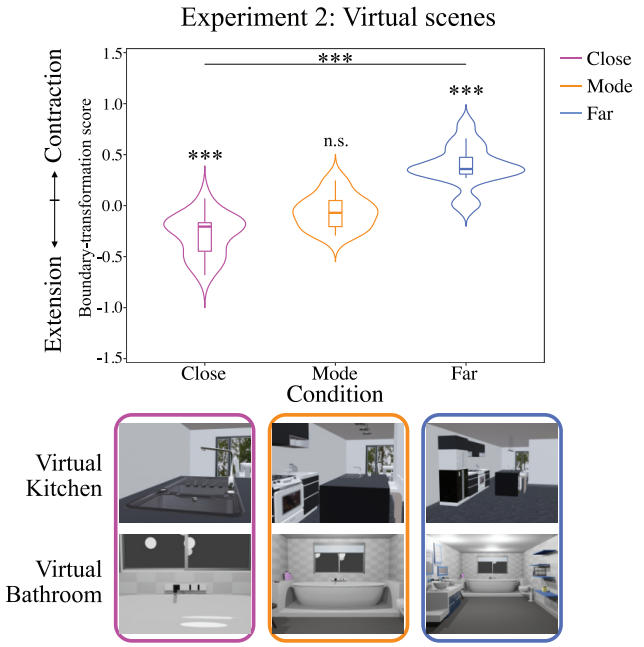
Using these more tightly controlled virtual scenes, we found a similar pattern of results as in our natural-scene experiments, with viewpoint judgements biased toward high-probability views. We first verified that boundary transformation scores in this experiment were consistent across split-half groups of participants ($\rho' = .94$, $p < .001$, Appendix Figure A2B). Next, we performed analyses across both items and participants, which showed results consistent with the natural scene experiments (see Figure 5). A one-way ANOVA on the item-level boundary transformation scores showed a significant effect of depth condition, $F(2, 45) = 50.84$, $p < .001$, $\eta^2 = .69$. Post hoc t tests with Bonferroni correction showed that the boundary transformation scores in both the Close condition ($M = -.29$) and Far condition ($M = .39$) differed significantly from zero and that scores in the Mode condition ($M = -.06$) did not differ from zero (Close: $t[15] = -5.61$, $p < .001$, $d = 1.40$, $BF_{10} = 5.25e2$; Far: $t[15] = 7.55$, $p < .001$, $d = 1.89$, BF = 1.06e4; Mode: $t[15] = -1.49$, $p = .473$, $d = .37$, $BF_{10} = .64$). A one-way ANOVA with Greenhouse-Geisser correction on the participant-level scores also showed a significant effect of depth condition, $F(1.36, 114.13) = 119.06$, $p < .001$, $\eta_p^2 = .59$. Post hoc t tests with Bonferroni correction again showed that the boundary transformation scores in both the Close condition ($M = -.29$) and Far condition ($M = .39$) differed significantly from zero, and that the scores in the Mode condition ($M = -.06$) did not differ from zero (Close: $t[84] = -7.00$, $p < .001$, $d = .76$, $BF_{10} = 1.72e7$; Far: $t[84] = 9.62$, $p < .001$, $d = 1.04$, $BF_{10} = 2.00e12$; Mode: $t[84] = -1.77$, $p = .242$, $d = .19$, $BF_{10} = .53$). Together, these findings reveal that differential boundary-transformation effects can be observed for images of the very same environment taken at different viewing distances. These effects are consistent with our prediction that scene memories are biased toward high-probability views in depth.

General Discussion

In this study, we developed a novel approach to quantify the statistical probabilities of scene views at varying depths in a large set of natural images, and we asked whether these view probabilities might explain boundary transformations in scene memory. Combining natural image statistics and behavioral data, we demonstrated that

Figure 5

Boundary-Transformation Effects on Virtual Scenes Revealed Biases Toward the Statistically Probable Image Depths of Natural Scenes



Note. Results from Experiment 2 using a more tightly controlled set of virtual scenes replicated the pattern of boundary transformation effects observed in Experiments 1a and 1b, with boundary extension for Close scenes, boundary contraction for Far scenes, and no consistent bias for Mode scenes. Plotting conventions for the violin plot are the same as in Figure 3. Example images are shown at the bottom of the figure. Asterisks above the line indicate significance for ANOVA, and those above each violin indicate significance for a two-tailed *t* test against zero. n.s. = not significant ($p > .05$). See the online article for the color version of this figure.
*** $p < .001$.

boundary extension and contraction can be explained by a single underlying mechanism that biases scene representations toward high-probability views. Earlier work speculated that such an inferential process could underlie memory biases but did not quantify the relevant statistics needed to test this idea (Bainbridge & Baker, 2020a; Intraub et al., 1992; Konkle & Oliva, 2007; McDunn et al., 2016). Our work is thus the first demonstration of the crucial link between the behavioral patterns of boundary transformation and the natural statistics of scene depth. These results situate boundary transformation within the broader framework of inferential cognitive processes that serve to increase the accuracy of noisy or incomplete representations by leveraging prior knowledge of the world (Griffiths et al., 2008; Kersten et al., 2004; Knill & Richards, 1996).

Importantly, there was no guarantee that the statistical regularities of image depth would be related to boundary transformation as we hypothesized. We could have alternatively found that depth probabilities were uninformative for predicting patterns of boundary extension and contraction or that the relationship between depth probabilities and boundary transformation was inconsistent with our theory. For example, a plausible alternative is that the highest probability views are the most familiar, making them ideal for eliciting strong anticipatory scene representations and, thus, strong boundary extension effects (Hubbard et al., 2010; Intraub, 2002). In contrast, we found no evidence of boundary extension (or contraction) for the highest probability views. Rather, we found that boundary transformation was far more likely to occur for lower probability views. Thus, our findings are the first to show that the effects of boundary extension and contraction are consistent with a memory bias for high-probability views based on the statistical regularities of image depths in natural scenes.

In addition, our work provides a systematic, quantitative approach for addressing a longstanding question about the role of prototypical views in boundary transformation. Researchers have long speculated about a possible mechanism that biases scene representations toward prototypical views (Bainbridge & Baker, 2020a; Intraub et al., 1992; Konkle & Oliva, 2007; McDunn et al., 2016). However, there have been no consistent or specific definitions in the literature of what constitutes a prototypical view, nor have there been any explicit quantitative models of this theory. Some have argued that the relevant prototypes are formed from the average of the stimuli that participants observe in an experiment (Intraub, 2020; McDunn et al., 2016), but such a theory cannot account for findings that show consistent image-level boundary transformation effects even when participants have viewed only a single experimental image (Bainbridge & Baker, 2020a, 2020b). Others have described prototypical views as those that are encountered most often in real-world scenes (Bainbridge & Baker, 2020a; Konkle & Oliva, 2007), but such proposals have been speculative thus far. Our statistical analysis of image depth is thus the first effort to directly quantify the natural scene statistics that underlie boundary transformation.

Although our findings demonstrated a strong relationship between image depth and boundary transformation, we could not perfectly control other nondepth properties. We attempted to mitigate the possibility that other nondepth factors were responsible for driving boundary transformation effects by conducting follow-up studies with more tightly controlled virtual scenes. Furthermore, recent work from our group has used visual illusions with natural scenes to demonstrate that depth, independent of other low- and high-level scene properties, plays a causal role in modulating boundary transformation (Hafri et al., in press). Thus, we suggest that the most parsimonious explanation for our findings is that boundary transformation relates to high-probability views in depth per se (and not other factors correlated with depth). This interpretation is consistent with other work using depth manipulations in virtual scenes or post hoc analyses of the scene properties that correlate with boundary transformation (Bainbridge & Baker, 2020a; Bertamini et al., 2005; Park et al., 2021).

There remain several exciting questions for future work. For example, in our studies, the statistical regularities that we quantified were based on scene views from photographs taken by researchers navigating through real-world environments. It is unknown whether these effects would generalize to views encountered in more naturalistic navigational behaviors, including in outdoor environments, which were not present in the image database that we analyzed when quantifying the statistical regularities of image depth. These ideas could be tested in future work by modeling the statistics of photographs of real-world environments and comparing these with the statistics of scene views encountered during natural behaviors in the same environments, perhaps by using head-mounted cameras to capture such views (Shankar et al., 2021).

Our discovery also raises questions of how statistical priors for scene views emerge in the first place. For example, future work may examine whether the effects of priors for depth are general or specific to different categories of scenes (e.g., office, auditorium). Our stimulus set did not allow us to definitively answer this question. Although we did find some suggestive evidence of at least some role for category-specific effects (see Figure 4), future work is needed to provide further support for such effects and to examine the conditions under which they arise. For example, they may depend on being able to recognize an image as belonging to a specific category in the first place, which for many close-up scenes, could prove difficult, as can be seen in Appendix Figure A1; in such cases, perhaps more general depth priors are the primary drivers of boundary-transformation effects. If it does turn out that category-specific priors have an influence on boundary transformation, it raises interesting questions about the role of individual experiences in developing such priors.

A related and important question is whether the relevant scene priors (either general or category-specific) develop from a lifetime of experience with views of naturalistic environments or whether they can be readily modified based on recent experiences. In other words, what kind of statistical learning is necessary to causally modulate boundary transformation? To directly test the causal role of recently learned priors, researchers could conduct a training study in which participants are exposed to scenes with experimentally controlled statistical distributions of views. One could then ask whether altering the statistical distributions of views induces changes in boundary transformation effects. We suspect that extensive training may be needed to override any potential preexisting priors that are learned over a lifetime of experience, as is generally found for other kinds of training studies (e.g., Cox et al., 2005; Op de Beeck et al., 2006). Nevertheless, these questions point to exciting future directions for better understanding how the statistics of visual experience influence scene representations in memory more generally.

Although boundary extension was discovered more than 30 years ago (Intraub & Richardson, 1989), the mechanisms underlying these memory distortions have remained elusive (Bainbridge & Baker, 2020a). Our findings show that boundary extension and contraction can be understood together as a memory bias for high-probability views in natural images. This suggests that boundary transformation results from a general computational principle for leveraging the statistical probabilities of experience to boost the accuracy of mental representations.

References

Bainbridge, W. A., & Baker, C. I. (2020a). Boundaries extend and contract in scene memory depending on image properties. *Current Biology*, *30*(3), 537–543.e3. https://doi.org/10.1016/j.cub.2019.12.004

Bainbridge, W. A., & Baker, C. I. (2020b). Reply to Intraub. *Current Biology*, *30*(24), R1465–R1466. https://doi.org/10.1016/j.cub.2020.10.032

Bertamini, M., Jones, L. A., Spooner, A., & Hecht, H. (2005). Boundary extension: The role of magnification, object size, context, and binocular information. *Journal of Experimental Psychology: Human Perception and Performance*, *31*(6), 1288–1307. https://doi.org/10.1037/0096-1523.31.6.1288

Cox, D. D., Meier, P., Oertelt, N., & DiCarlo, J. J. (2005). 'Breaking' position-invariant object recognition. *Nature Neuroscience*, *8*(9), 1145–1147. https://doi.org/10.1038/nn1519

Griffiths, T. L., Kemp, C., & Tenenbaum, J. B. (2008). Bayesian models of cognition. In R. Sun (Ed.), *The Cambridge Handbook of Computational Psychology* (pp. 59–100). Cambridge University Press. https://doi.org/10.1017/CBO9780511816772.006

Hafri, A., Wadhwa, S., & Bonner, M. F. (in press). Perceived distance alters memory for scene boundaries. *Psychological Science*. https://doi.org/10.31234/osf.io/hy3qs

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

Intraub, H. (2002). Anticipatory spatial representation of natural scenes: Momentum without movement? *Visual Cognition*, *9*(1–2), 93–119. https://doi.org/10.1080/13506280143000340

Intraub, H. (2010). Rethinking scene perception: A multisource model. *Psychology of Learning and Motivation*, *52*, 231–264. https://doi.org/10.1016/S0079-7421(10)52006-1

Intraub, H. (2020). Searching for boundary extension. *Current Biology*, *30*(24), R1463–R1464. https://doi.org/10.1016/j.cub.2020.10.031

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. https://doi.org/10.1111/j.1467-9280.2008.02192.x

Intraub, H., & Richardson, M. (1989). Wide-angle memories of close-up scenes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(2), 179–187. https://doi.org/10.1037/0278-7393.15.2.179

Intraub, H., Bender, R. S., & Mangels, J. A. (1992). Looking at pictures but remembering scenes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(1), 180–191. https://doi.org/10.1037/0278-7393.18.1.180

Kersten, D., Mamassian, P., & Yuille, A. (2004). Object perception as Bayesian inference. *Annual Review of Psychology*, *55*(1), 271–304. https://doi.org/10.1146/annurev.psych.55.090902.142005

Knill, D. C., & Richards, W. (1996). *Perception as Bayesian inference*. Cambridge University Press. https://doi.org/10.1017/CBO9780511984037

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*, 6.

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. https://doi.org/10.1080/13506285.2016.1274810

Op de Beeck, H. P., Baker, C. I., DiCarlo, J. J., & Kanwisher, N. G. (2006). Discrimination training alters object representations in human extrastriate cortex. *The Journal of Neuroscience*, *26*(50), 13025–13036. https://doi.org/10.1523/JNEUROSCI.2481-06.2006

Park, J., Josephs, E. L., & Konkle, T. (2021, January 26). *Systematic transition from boundary extension to contraction along an object-to-scene continuum*. https://doi.org/10.31234/osf.io/84exs

Petzschner, F. H., Glasauer, S., & Stephan, K. E. (2015). A Bayesian perspective on magnitude estimation. *Trends in Cognitive Sciences*, *19*(5), 285–293. https://doi.org/10.1016/j.tics.2015.03.002

Shankar, B., Sinnott, C., Binaee, K., Lescroart, M. D., & MacNeilage, P. (2021). Ergonomic Design Development of the Visual Experience Database Headset. *ACM Symposium on Eye Tracking Research and Applications*, 1–4. https://doi.org/10.1145/3450341.3458487

Song, S., Lichtenberg, S. P., & Xiao, J. (2015). SUN RGB-D: A RGB-D scene understanding benchmark suite. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 567–576). https://doi.org/10.1109/CVPR.2015.7298655

Appendix

Supplementary Figures

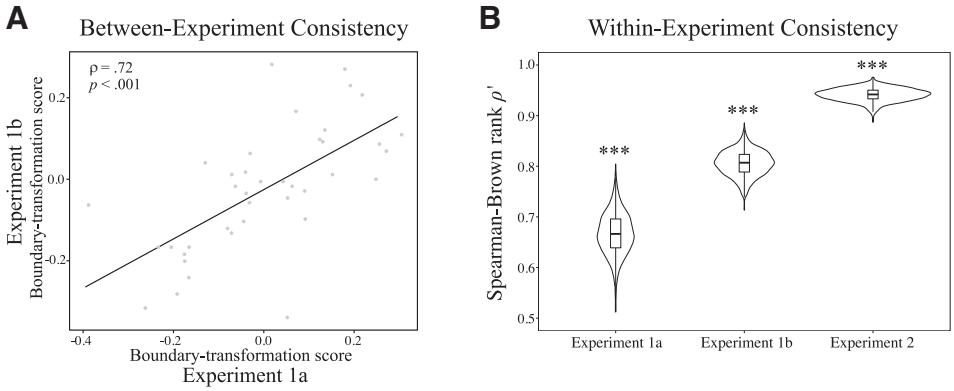
Figure A1
Examples of Natural-Scene Stimuli

Note. Example images are shown for each of the 15 natural scene categories in the Close, Mode, and Far conditions. See the online article for the color version of this figure.

(*Appendix continue*)

Figure A2

Boundary Transformation Effects Are Consistent Within and Between Experiments



Note. (A) Between-experiment consistency was assessed for the two natural-scene experiments. Based on Spearman's rank correlation, these category-wise boundary transformation scores were remarkably consistent across the two experiments, which involved different sets of stimuli. (B) Within-experiment consistency among participants was assessed for all three experiments by comparing the image-wise boundary-transformation scores across split-half groups of participants over 1,000 iterations. These violin plots show distributions of split-half correlations across the 1,000 iterations. *p* values were computed with permutation tests, in which null distributions of correlation values were calculated using the same procedure and the same 1,000 split-half groups of participants except that the image order in one group of participants was randomized. These analyses confirmed that boundary transformation scores were reliable across participants in all three experiments. Asterisks above violins indicate significance through permutation tests.
*** $p < .001$.

(*Appendix continue*)

Figure A3
Examples of Virtual-Scene Stimuli

Note. Example images are shown for each of the nine scene categories in the Close, Mode, and Far conditions. See the online article for the color version of this figure.

Received March 24, 2022
Revision received June 14, 2022
Accepted June 17, 2022 ■