Article
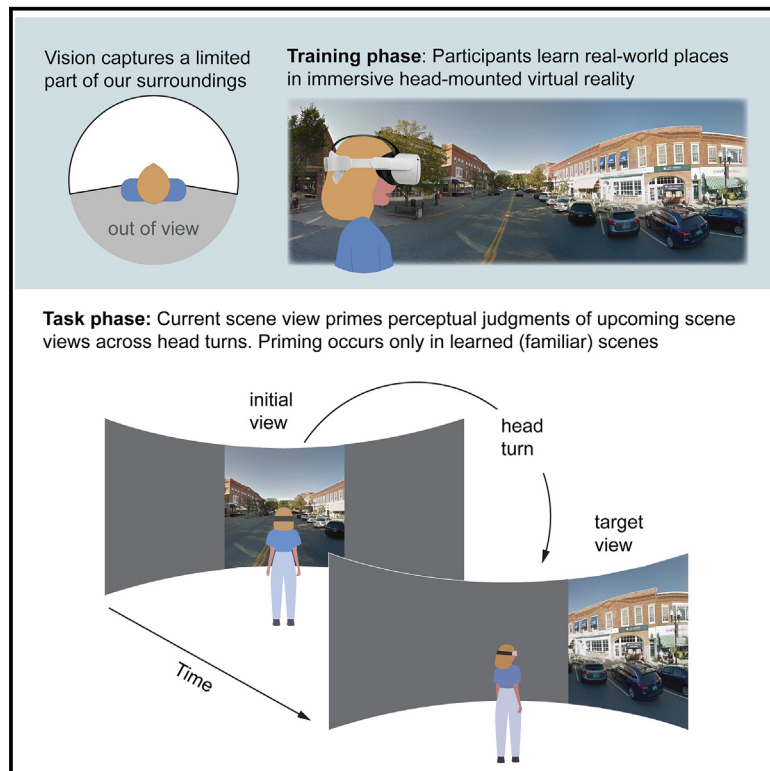
Memory-based predictions prime perceptual judgments across head turns in immersive, real-world scenes

Graphical abstract

Authors

Anna Mynick, Adam Steel,
Adithi Jayaraman, Thomas L. Botch,
Allie Burrows, Caroline E. Robertson

Correspondence

annamynick@gmail.com (A.M.),
caroline.e.robertson@dartmouth.edu
(C.E.R.)

In brief

Our field of view only captures part of our environment, but we experience our environment as a unified whole. How is this accomplished? Using immersive VR, Mynick et al. show that memory for the current environment primes perceptual judgments of upcoming views across head turns, supporting efficient perception in panoramic environments.

Highlights

- As we look around a place, perceptual judgments are primed across head turns

- Memory for the current place is required for priming across head turns

- Action plans dictate primed content across head turns

CellPress

Current Biology

Article

Memory-based predictions prime perceptual judgments across head turns in immersive, real-world scenes

Anna Mynick,[1,2,*] Adam Steel,[1] Adithi Jayaraman,[1] Thomas L. Botch,[1] Allie Burrows,[1] and Caroline E. Robertson[1,*]
[1]Department of Psychological and Brain Sciences, Dartmouth College, 3 Maynard Street, Hanover, NH 03755, USA
[2]Lead contact
*Correspondence: annamynick@gmail.com (A.M.), caroline.e.robertson@dartmouth.edu (C.E.R.)
https://doi.org/10.1016/j.cub.2024.11.024

SUMMARY

Each view of our environment captures only a subset of our immersive surroundings. Yet, our visual experience feels seamless. A puzzle for human neuroscience is to determine what cognitive mechanisms enable us to overcome our limited field of view and efficiently anticipate new views as we sample our visual surroundings. Here, we tested whether memory-based predictions of upcoming scene views facilitate efficient perceptual judgments across head turns. We tested this hypothesis using immersive, head-mounted virtual reality (VR). After learning a set of immersive real-world environments, participants (n = 101 across 4 experiments) were briefly primed with a single view from a studied environment and then turned left or right to make a perceptual judgment about an adjacent scene view. We found that participants' perceptual judgments were faster when they were primed with images from the same (vs. neutral or different) environments. Importantly, priming required memory: it only occurred in learned (vs. novel) environments, where the link between adjacent scene views was known. Further, consistent with a role in supporting active vision, priming only occurred in the direction of planned head turns and only benefited judgments for scene views presented in their learned spatiotopic positions. Taken together, we propose that memory-based predictions facilitate rapid perception across large-scale visual actions, such as head and body movements, and may be critical for efficient behavior in complex immersive environments.

INTRODUCTION

Humans constantly make predictions to support ongoing behavior: we predict what will be inside our mailbox, which people will be in the room we are entering, and whether there is time to get through a yellow light. These predictions are often shaped by our memories of past experiences.[1] In traditional studies of memory-based predictions, participants learn associations between stimuli presented sequentially on a screen and are trained to predict upcoming target images (e.g., expect C, given A and B).[2–7] Once learned, these associations facilitate sensory processing of upcoming stimuli, speeding detection[8,9] and improving recognition of anticipated percepts.[10,11] However, despite their presumed importance in everyday cognition, the content and form of memory-based predictions during naturalistic, active vision are unclear. As we look around the world, exchanging visual content in the current view for content in the next, what information is predicted across head turns?

Several lines of evidence suggest that the visual system predicts the perceptual consequences of small-scale visual actions (i.e., saccades). During saccades, associations between foveal and peripheral percepts support perceptual continuity. For instance, before a saccade to a target stimulus, the pre-saccadic center of gaze already shows enhanced sensitivity to the post-saccadic target's stimulus features, linking processing of the current view to the upcoming one.[12] Memory for a previously foveal percept also influences peripheral vision. For example, past work has shown that the percept of a stimulus in the periphery can become biased to resemble the same stimulus' foveal percept.[13,14] These behavioral findings align closely with neural recordings. For example, cells in monkey retinotopic cortex anticipate the visual consequences of saccades by preemptively responding to objects that will fall into their receptive fields following a saccade.[15–18] Compatible evidence for this phenomenon—termed "predictive remapping"[15]—has also been identified in human neuroimaging[19–21] and behavior.[22–25] So, the visual system clearly predicts the consequences of small-scale visual actions—i.e., saccades—but what about larger-scale behaviors? Prediction during saccades is relatively straightforward because the visual information being predicted is, of course, already available to draw from in the visual periphery. In contrast, an open question is how predictive vision generalizes to large-scale visual actions like head and body movements, where upcoming visual information is fully out of view.

Two solutions could plausibly address how the brain generates predictions across head turns in immersive, real-world contexts. First, schema-level knowledge about the type of environment could be used to extrapolate visual content beyond the current field of view. Scene extrapolation has been proposed as a mechanism supporting the phenomenon of boundary
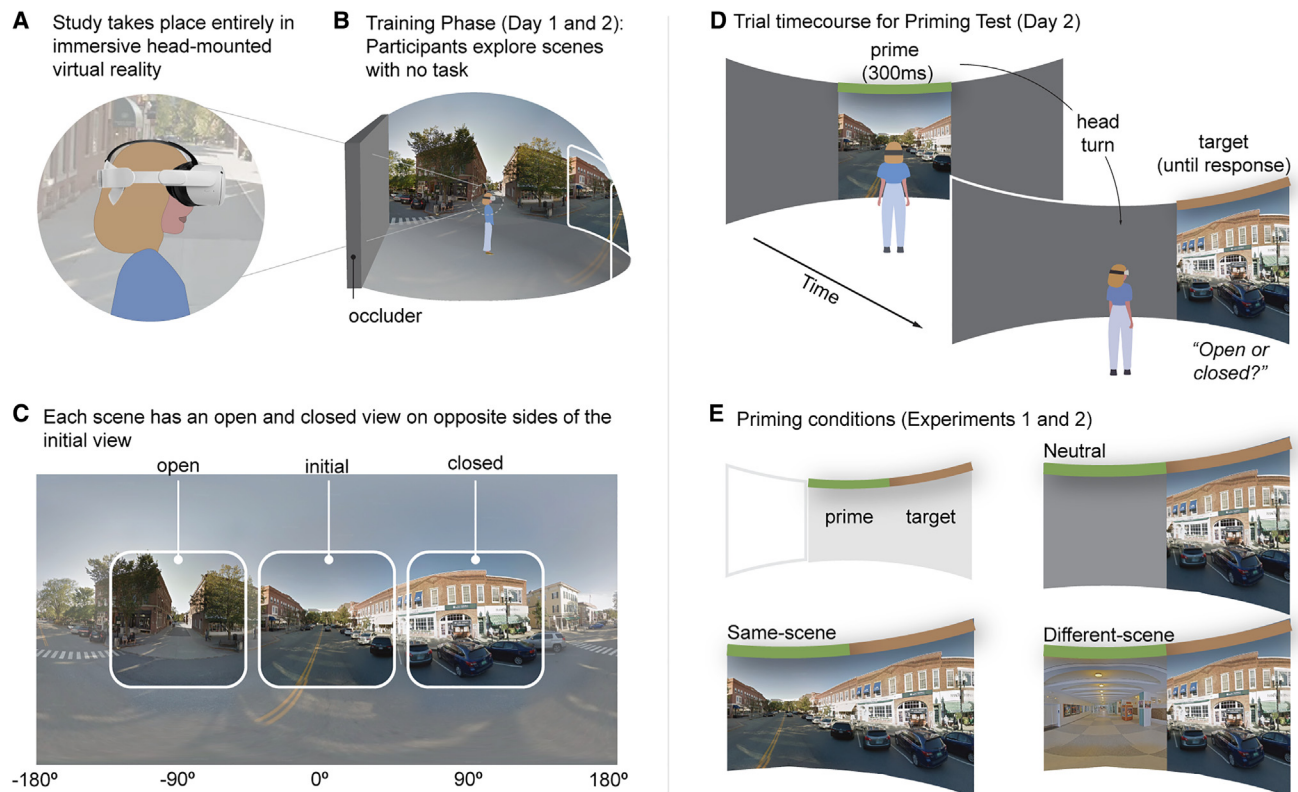
Figure 1. Experimental paradigm
(A) Participants were fitted with virtual reality (VR) headsets, which were used throughout the study.
(B) On both days of the training phase, participants studied each real-world scene in immersive VR for 20 s, instructed to simply "look around like you normally would."
(C) Each scene had one open and one closed view on opposing sides of the initial facing-direction in the scene, spaced 180° apart on the horizontal axis.
(D) On day 2, participants performed a priming test in which they viewed a 300 ms prime image before turning left or right toward a target image (open or closed view) to make a speeded perceptual (open/closed) judgment. Targets were presented until a response was made.
(E) Prime images in experiments 1 and 2 either depicted a continuous view adjacent to the target from the initial direction (same-scene), a blank gray rectangle (neutral), or an image taken from the initial direction of a different scene (different-scene).
See also Figure S1 for full set of scene stimuli, Figure S2 for familiarity responses, Figure S3 for perceptual judgment accuracy, and Videos S1, S2, and S3 for screen videos of training phase and priming test.

extension, whereby participants falsely report having seen a zoomed-out version of a previously viewed scene image[26–29] (but see Bainbridge and Baker,[30] Gandolfo et al.,[31] and Hafri et al.[32]). Second, memory of the broader visuospatial environment could supply specific predictions about the content and layout of upcoming views that will result from visual actions. Indeed, recent work has shown that views from the same environment become associated in the brain, which could support this type of behavioral prediction.[33,34]

Here, we sought to understand how memory supports naturalistic scene perception across large-scale visual actions in immersive environments. To do this, we developed an experimental paradigm that used head-mounted virtual reality (VR) to examine memory-based predictions in immersive, real-world scenes. After learning a set of immersive real-world scenes, participants were primed with a single scene view before head-turning to see another view from the studied scenes and making an ecologically relevant perceptual judgment ("could I walk into this space?").

To preview our results, we found that memory-based predictions are generated for upcoming scene views across head turns

(experiment 1) and that these predictions depend on memory for out-of-sight views of the current scene (experiment 2). Crucially, these predictions appear to occur in service of action plans (experiments 3 and 4), priming responses to content that is congruent with the visual outcome of an intended head turn. Together, our results suggest a role for memory-based predictions in facilitating rapid perception as we sample the visual world around us.

RESULTS

Upcoming scene views are primed across head turns
We first asked whether memory-based predictions are generated for upcoming scene views across head turns in immersive environments (experiment 1). In the training phase, participants ($n = 18$) learned a set of immersive real-world scenes in VR (Figure 1A). To facilitate faster learning, scenes depicted familiar locations around the local university (experiment 1 familiarity per participant, quantified by percent "yes" responses to the prompt "are you familiar with this place in real life? [yes/no/unsure]": M = 85.6%,

SD = 11.63%) (Figures S1 and S2). Scenes fully surrounded participants, so that they needed to use head turns to see a scene completely (Figure1B). On each trial (20 s), participants were simply told to naturally explore each scene (i.e., "look around like you normally would"). Each scene featured an open view (e.g., a view of an open parking lot) and a closed view (e.g., a view of a building's façade) on either side of an initial facing direction (Figure 1C). Importantly, participants could accurately recall the relative spatial position of each scene view, quantified by recall accuracy for whether a view had appeared on the left, right, or center of a scene during the training phase (explicit memory test, M = 93.70 ± 5.80 SD; t(17) = 44.17, $p < 0.001$, 95% confidence interval [CI] [90.81, 96.58]) (Figure S3).

After the training phase, we assessed participants' memory-based predictions in the priming test. On each trial, participants saw a full-field scene image ("prime") from a studied scene. Following the prime image, participants turned approximately 90° left or right to make a perceptual judgment (open or closed) on a second scene view ("target") that appeared in their periphery (Figure 1D). Participants were accurate at making this judgment (Figure S3C; Table 2). Primes either depicted: (1) a view from the same scene (same-scene prime), (2) a blank gray rectangle (neutral prime), or (3) a view from a different scene (different-scene prime) (Figure 1E). We considered memory retrieved in response to the prime image the "predicted content," and we assessed the behavioral impact of this prediction by comparing median response time (Figure 2A) across the priming conditions. We hypothesized that if participants generate memory-based predictions for upcoming scene views, response times would be faster after same-scene primes compared with neutral or different-scene primes. We compared the differences across conditions using linear mixed effects (LME) models with priming condition (same-scene/neutral/different-scene) as a fixed effect.

Our results reveal evidence for memory-based predictions across head turns. A LME model (fixed effect: priming condition; random effects: participant and scene) showed a significant effect of priming condition on response time (F(2, 1,908.7) = 33.78, $p < 0.001$, $\eta_p^2 = 0.03$) (Figure 2B; Table S1). Critically, perceptual judgments were faster after same-scene primes compared with neutral primes (t(17) = −2.77, $p = 0.01 \times 3$ tests = $p_{corr} = 0.04$, d = −0.43, 95% CI [−0.07, −0.01]) and were slower after different-scene primes compared with neutral primes (t(17) = −3.37, $p_{corr} = 0.01$, d = 0.57, 95% CI [−0.08, −0.02]). This priming effect (same-scene < neutral) also replicated across three pilot datasets (Figure S4). Interestingly, priming was also observed in scenes that were reported as unfamiliar during the day 1 training phase (t(102) = −2.79, $p = 0.006$, d = −0.27, 95% CI [−0.08, −0.01]) (Figure S2C), suggesting that scenes learned in virtual environments can support memory-based prediction without real-world scene experience. Taken together, these results demonstrate that upcoming scene views are primed across head turns in familiar environments.

Motor response preparation does not account for priming effect

This priming effect seems to suggest that predictions are generated based on memory for the visual content of the upcoming

scene view. An alternative interpretation, however, is that participants learned to associate each prime view with a motor response (e.g., "if I turn left, I'll press the 'open' button, if I turn right, the I'll press the 'closed button'") without actually predicting the visual content of the upcoming view. To rule out this possibility, we examined trials from the different-scene condition, in which the prime and target scene either resulted in "motor-matched" trials because they had matched layouts (e.g., were both open on the left, closed on the right) or "motor-mismatched" trials because they had mismatched layouts (e.g., the prime was closed on the right, the target was open on the right) (Figure 2C). If participants had learned to associate each prime view with a motor response plan instead of generating a visual prediction, motor-matched primes should evoke the correct motor plan (albeit the wrong visual prediction), resulting in faster response times for motor-matched than motor-mismatched trials. Instead, a paired t test revealed no difference in response time between motor-matched and -mismatched trials (t(17) = 1.52, $p = 0.146$, 95% CI [−0.01, 0.06]) (Figure 2D). Indeed, the data in fact trend in the opposite direction, with numerically faster responses in motor-mismatched trials. Taken together, this suggests that information about upcoming visual content—rather than just a motor response—was primed across head turns.

Priming relies on memory for immersive environments

We next tested whether priming across head turns requires memory for the specific scene or whether predictive information arises from scene extrapolation (akin to boundary extension).[27] To do this, in experiment 2, we invited a new set of participants (n = 20) to perform the same priming test as in experiment 1, but this time on a set of novel, unfamiliar scenes with no training phase (Figure S5 for scene images; Table 1). We then calculated the priming effect (same-scene vs. neutral vs. different-scene) in experiment 2 (unfamiliar scenes) and compared the extent of the priming effect with the effect in experiment 1 (familiar scenes). Experiment 2 participants were accurate at making open/closed judgments (Figure S3C; Table 2). To measure the extent of priming, we compared response times across the three conditions using a LME model. This analysis revealed that there was no difference in response times across priming conditions in experiment 2 alone (LME model: F(2, 1,889.9) = 0.70, $p = 0.49$, $\eta_p^2 = 0.0007$) (Figure 2E; Table S1), suggesting that no predictive information was available to speed up perceptual judgments. Next, we compared response times for unfamiliar scenes with those for familiar scenes (i.e., response times from experiment 1), using a LME model with priming condition and experiment (experiment 1/experiment 2) as fixed effects (participant and scene as random effects). Overall, response times were slower for unfamiliar scenes (main effect of experiment: F(1, 42.7) = 15.72, $p < 0.001$, $\eta_p^2 = 0.27$). Critically, there was a significant interaction between experiment and priming condition (F(2, 3,861.5) = 19.012, $p < 0.001$, $\eta_p^2 = 0.001$), indicating that familiarity with a scene is required for priming across head turns. Together, our results indicate that behavioral priming across head turns requires memory-based prediction and does not arise from scene extrapolation.
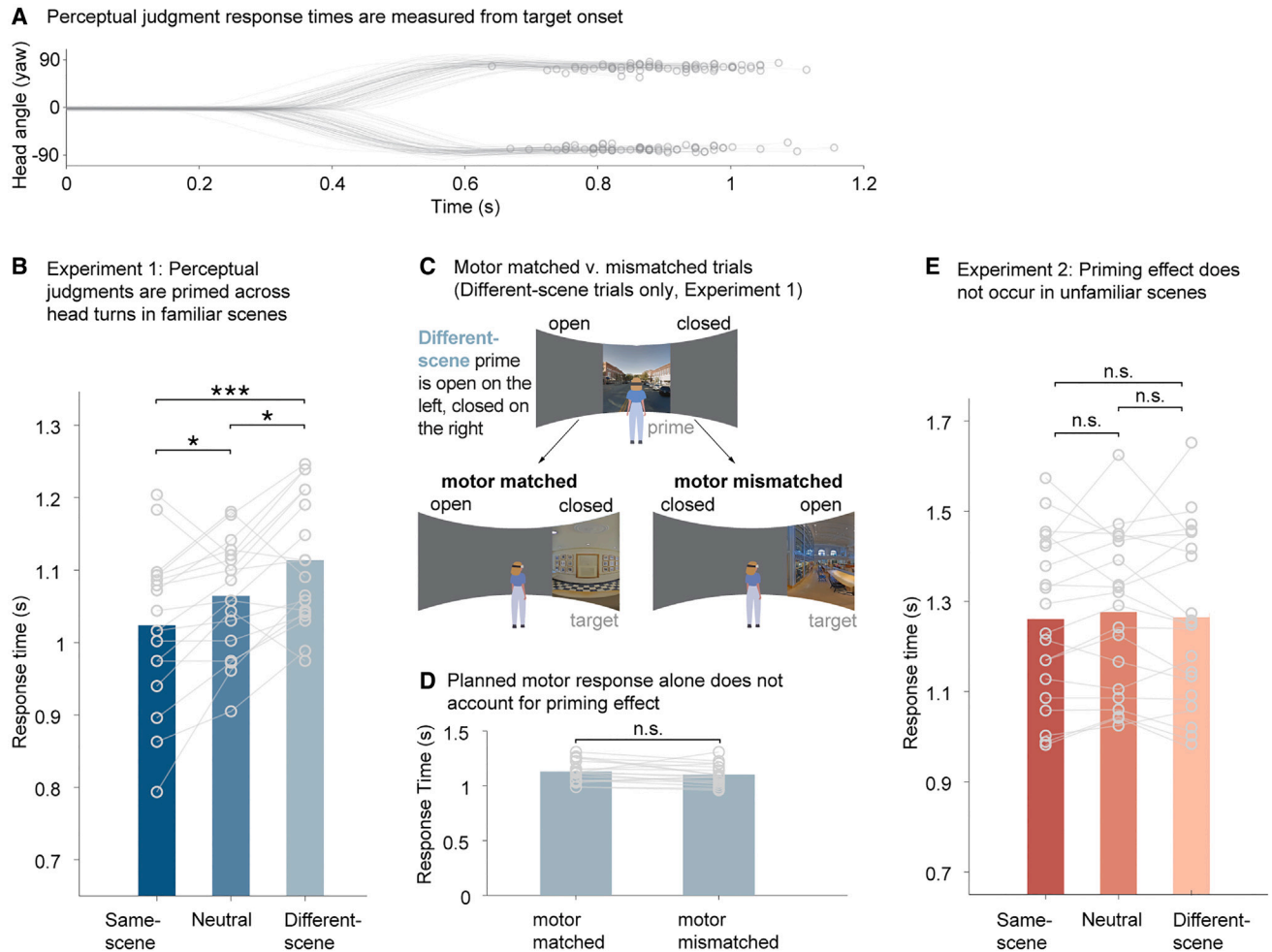
A Perceptual judgment response times are measured from target onset

B Experiment 1: Perceptual judgments are primed across head turns in familiar scenes

C Motor matched v. mismatched trials (Different-scene trials only, Experiment 1)

Different-scene prime is open on the left, closed on the right

open   closed

prime

motor matched
open   closed

motor mismatched
closed   open

target

target

D Planned motor response alone does not account for priming effect

n.s.

motor matched   motor mismatched

E Experiment 2: Priming effect does not occur in unfamiliar scenes

n.s.

n.s.   n.s.

Same-scene   Neutral   Different-scene

Figure 2. Same-scene primes facilitate faster perceptual judgments in familiar (but not unfamiliar) scenes

(A) Example participant's data depicting head angle on the horizontal plane across a trial. Each line depicts head angle on the horizontal plane (yaw) for one trial from target onset. Each open circle marks the (time, head angle) of the participant at the time of the perceptual judgment.

(B) We compared median response times for perceptual judgments across priming conditions in experiment 1 ($n = 18$). Compared with neutral primes, same-scene primes quickened response times and different-scene primes slowed response times. Connected points represent the same participant across conditions.

(C) Different-scene trials (experiment 1) either contained (1) a prime and target with matching layouts (e.g., both open on the left, closed on the right), in which the prime would evoke the correct motor plan (motor matched), or (2) mismatched layouts (e.g., prime open on the left, target open on the right) in which the prime would evoke the wrong motor plan (motor mismatched).

(D) Response times in different-scene motor-matched trials and motor-mismatched trials do not differ. Each participant's median response time was calculated for motor-matched vs. motor-mismatched trials (paired t test). Connected points represent the same participant across bars.

(E) There was no difference in median response times for perceptual judgments across priming conditions in experiment 2 ($n = 20$), which contained unfamiliar scenes. Connected points represent the same participant across conditions.

For all plots, $^*p \leq 0.05$, $^{**}p \leq 0.01$, $^{***}p \leq 0.001$ difference between median response time for indicated priming conditions (Bonferroni-corrected paired t tests); n.s., not significant.

See also Figure S2 for familiarity responses, Figure S3 and Table S3 for priming test accuracy, Figure S4 for replication of experiment 1 priming effect (same-scene vs. neutral), Figure S5 for experiment 2 scene stimuli, Tables S1 and S2 for additional analysis, and Video S1 for screen videos of training phase and priming test.

Action plans bias direction of priming

So far, the results of experiments 1 and 2 suggest that memory-based predictions support efficient scene perception across head turns. In both of these experiments, an *exogenous* cue (a target appearing in the periphery) alerted participants to where they should look next. But most often in real-world contexts, an *endogenous* action plan directs where we look (e.g., I head turn right to see whether I can safely change lanes in traffic).[35,36] Do *endogenous* action plans exaggerate the extent of memory-based predictions across head turns?

To test this, in experiment 3, we asked whether predictions are biased in the direction of intended action plans or whether predictions are generated for all scene views surrounding the current one, regardless of action plans. Participants ($n = 26$)

Table 1. Overview of experiments

–	# Days	Scenes	# Priming test trials	Priming test conditions	Condition split (%)	# Scene view repetitions by condition	Neutral type	Explicit memory test	# Participants collected	# Participants included
Experiment 1	2	local university	144	same-scene	50	2	blank gray	yes	26	18
				neutral	25	1				
				different-scene	25	1				
Experiment 2	1	unfamiliar locations	144	same-scene	50	1	blank gray	no	24	20
				neutral	25	<1				
				different-scene	25	<1				
Experiment 3	2	local university	216	same-scene, valid arrow	50	3	scrambled	no	46	26
				neutral, valid arrow	16.7	1				
				same-scene, invalid arrow	16.7	1				
				neutral, invalid arrow	16.7	1				
Experiment 4	2	local university	144	same-scene spatially congruent	50	2	blank gray	yes	44	37
				neutral	25	1				
				same-scene spatially displaced	25	1				
Pilot A	2	local university	108	same-scene	66.6	2	blank gray	yes	23	18
				neutral	33.3	1				
Pilot B	2	local university	108	same-scene	66.6	2	scrambled	yes	31	21
				neutral	33.3	1				
Pilot C	2	local university	108	same-scene	66.6	2	scrambled	yes	28	22
				neutral	33.3	1				

completed a version of our priming test using the same familiar stimulus set as in experiment 1, in which a brief left or right arrow (300 ms) preceded the prime images (same-scene or neutral) (Figure 3A), indicating a direction to plan a head turn in. Arrows correctly indicated the upcoming target location ("valid arrow") on two-thirds of trials and were incorrect ("invalid arrow") in the remaining one-third of trials. If predictions serve intended action plans, a planned head turn should strengthen the extent of priming in the intended—but not unintended—direction.

Our results were consistent with this hypothesis. As in previous experiments, participants were accurate at making open/closed judgments (Figure S3C; Tables 2 and S3). We observed both a main effect of priming condition (same-scene/neutral) (LME model: $F(1, 4,405.8) = 35.16$, $p < 0.001$, $\eta_p^2 = 0.008$) and arrow condition (valid/invalid) ($F(1, 4,406.7) = 64.98$, $p \leq 0.001$, $\eta_p^2 = 0.01$) on response time. Critically, the priming effect was stronger in the direction of intended head turns, evidenced by a significant interaction between priming condition (same-scene/neutral) and arrow condition (valid/invalid) ($F(1, 4,405.9) = 9.32$, $p = 0.002$, $\eta_p^2 = 0.002$). Post hoc tests indicated that same-scene primes resulted in faster response times than neutral primes in trials with valid arrows ($t(25) = -4.26$, $p_{corr} = 0.002$, d = -0.29, 95% CI [-0.07, -0.02]) but not invalid arrows ($t(25) = -0.65$, $p_{corr} = 1$, d = 0.07, 95% CI [-0.04, 0.02])

(Figure 3B; Figure S6). This finding suggests that memory-based predictions are coordinated with planned actions to facilitate efficient perceptual judgments in immersive environments.

Action plans dictate primed content

Finally, we turned to the content of the memory-based prediction—is the content of a memory-based prediction determined by an agent's planned actions (i.e., head-turn direction) within a spatial environment? Using the priming test structure established in experiments 1 and 2 (i.e., no arrow cue), in experiment 4 ($n = 37$), we introduced a new priming condition (the "same-scene spatially displaced" condition) in which the prime and target were drawn from the same scene but the target image depicted the view 180° opposite the expected one (e.g., the scene's left view, presented where the right view should appear in space) (Figure 4A). This created a conflict between an action (e.g., right head turn) and the predicted visual consequence of that action (e.g., left scene view instead of right scene view). If the content of memory-based predictions were determined by an agent's action in a spatial environment, then priming should remain intact for same-scene (spatially congruent) trials but should be disrupted for same-scene spatially displaced ones.

Consistent with our hypothesis that an agent's actions within an immersive environment determine the content of memory-based

Table 2. Overview of priming test accuracy (experiments 1–4)

–	Mean accuracy (SD) (%)	Repeated-measures ANOVA	sig	df	F	p	η_p^2	Post hoc pairwise t tests	Condition comparison	sig	df	t	p^a
Experiment 1	91.44 (4.88)	–	n.s.	(2,34)	0.92	0.41	0.05	–	–	–	–	–	–
Experiment 2	90.42 (4.34)	–	**	(2,38)	6.01	0.005	0.24	–	same-scene vs. neutral	n.s.	19	−0.72	1
									same-scene vs. different-scene	*	19	3.23	0.013
									neutral vs. different-scene	*	19	2.77	0.036
Experiment 3	88.43 (5.35)	–	n.s.	(1,25)	0.02	0.89	8e−5	–	–	–	–	–	–
Experiment 4	91.12 (0.05)	–	***	(1.65,59.31)	9.41	p < 0.001	0.21	–	same-scene vs. neutral	**	36	3.20	0.009
									same-scene spatially congruent vs. same-scene spatially displaced	n.s.	36	−1.19	0.73
									neutral vs. same-scene spatially displaced	**	36	−3.50	0.004

Results from repeated-measures ANOVA and post hoc t tests evaluating differences in priming test open/closed accuracy between priming conditions. Sig, significance for indicated statistical test; *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$. df, degree(s) of freedom.
^ap values after Bonferroni correction.

predictions, we found a significant effect of priming condition on response time (LME model: $F(2, 3,811.9) = 14.66$, $p < 0.001$, $\eta_p^2 = 0.008$) (Figure 4B; Table S1), whereby same-scene spatially congruent primes sped response time compared with neutral ones ($t(36) = −2.11$, $p = 0.04$, $d = −0.21$, 95% CI [−0.05, −0.001]), replicating the priming effect observed in experiments 1 and 3. Critically, same-scene spatially displaced trials elicited slower response times than same-scene spatially congruent trials ($t(36) = −2.14$, $p = 0.04$, $d = −0.28$, 95% CI [−0.07, −0.002]), suggesting that the conflict between visual action and visual outcome disrupted memory-based predictions across head turns. Taken together, these results argue for a mechanism by which predicted content may be mapped to the spatial scaffold of a scene based on an agents' action in that environment.

DISCUSSION

Here, we show that memory-based predictions of upcoming, out-of-sight scene views prime perception of those views across head turns to enable rapid, ecologically relevant perceptual judgments. This priming only occurs in familiar scenes, showing that it depends on memory for the out-of-sight content. Further, priming is skewed in the direction of planned head turns, underscoring its utility for active vision. Based on these results, we propose that memory-based predictions likely enable us to overcome our limited field of view by supporting efficient perception across large-scale visual actions.

Our work advances our understanding of how predictive processes influence perception across large-scale visual actions by showing that memory-based predictions are generated for upcoming scene views across head turns in immersive environments (experiments 1–4). Importantly, we identify evidence of upcoming scene view prediction in three separate experiments:

experiment 1, experiment 3 (valid trials), and experiment 4 showed faster responses to expected (same-scene) scene views than unexpected (different-scene) or unpredictable (neutral) ones. Previous studies of predictive vision have primarily used basic stimuli (e.g., oriented gratings or single objects) to examine small-scale visual actions like saccades within a single field of view. In these contexts, others have found evidence for predictive vision in foveal and peripheral vision.[12–17] For example, before making a saccade, the center of gaze becomes more sensitive to the visual features of the stimulus that the saccade will bring into view.[12] Likewise, across saccades, the appearances of peripheral objects become biased toward their associated foveal percepts, which is thought to smooth the transition between low-acuity peripheral vision and high-acuity foveal vision.[13,14] Interestingly, saccades are necessary to evoke this predictive bias—biased perception does not occur for percepts that become associated without a saccade,[13] which mirrors our finding that intended actions modulate the extent of priming (experiment 3). However, prediction across saccades differs fundamentally from predictions across head turns: across saccades, information in the periphery can be used to inform predictions,[13] whereas head turns often bring new information into the field of view and should therefore require additional information to inform predictions. Indeed, our results indicate that, in immersive contexts, memory supports predictions of upcoming scene views across head turns. Thus, an open question is the degree of overlap between the cognitive mechanisms underpinning predictions across saccades (operating within a field of view) vs. across large-scale visual actions like head turns (operating across many fields of view).

Our finding that priming occurs across head turns in familiar (but not unfamiliar) scenes suggests the existence of a memory-dependent cognitive mechanism that supports efficient
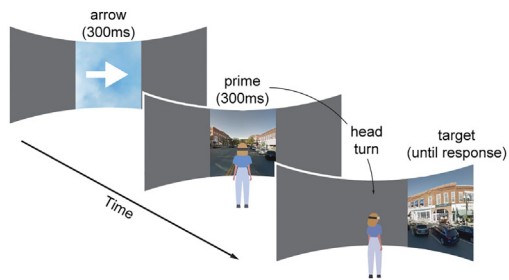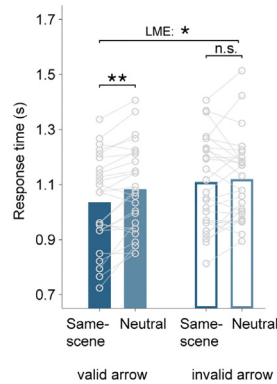
Figure 3. Priming is biased in the direction of intended head turns

(A) On each trial of experiment 3, participants saw an arrow indicating a direction to plan a head turn in, followed by a prime and target. Arrows were either valid (i.e., correctly indicated the location of the upcoming target) or invalid.

(B) Median response times for perceptual judgments across priming conditions in experiment 3 ($n = 26$) were only faster in same-scene compared with neutral primes when preceded by a valid arrow. Connected points represent the same participant across conditions. * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$ difference between median response time for indicated priming conditions (Bonferroni-corrected paired t tests unless otherwise noted). LME indicates the interaction between arrow condition and priming condition using a linear mixed effects model. See also Figure S6, Tables S1 and S2 for additional analysis, and Video S1 for screen videos of training phase and priming test.

visual perception across head turns. This memory-dependent mechanism likely differs fundamentally from two known phenomena that occur robustly in response to unfamiliar images—boundary extension and scene layout extrapolation[37–39]—which both indicate that out-of-sight information can be extrapolated from a limited field of view (but see Shafer-Skelton and Brady[40]). However, boundary extension and scene layout extrapolation are inherently different from predictions across large-scale visual actions like head turns: both operate only within a small spatial window around the scene.[27,37] Thus, our findings likely complement the cognitive phenomena of boundary extension or scene layout extrapolation, contributing instead to a growing body of literature showing that memory is leveraged to proactively support behavior.[8,9,41–45] We hypothesize that this cognitive mechanism, memory-based scene view prediction, could arise from associations in the brain between scene views sampled from the same environment during learning.[33,34]

The present work describes evidence for a predictive process in which *visual* predictions of upcoming scene views are generated to support active vision. Importantly, our results also argue against the alternative interpretation that prime images elicited simple *motor* predictions or *response biases* (e.g., "if I turn left, I'll press the 'open' button"). Using data from different-scene trials (experiment 1), we examined trials where the prime and target had matching spatial layouts, such that the prime would evoke the correct motor plan but the wrong visual prediction (motor-matched trials). Critically, these trials were no faster than trials with mismatched layouts (wrong motor plan, wrong visual prediction), indicating that preparing a motor response alone is not sufficient to account for the same-scene priming effect we observe across our experiments. Relatedly, though our decision to exclude trials where participants head-turned before the target presentation led to a relatively high exclusion rate in experiment 3 (36.5% of trials), this choice also bolsters the likelihood that the priming effect observed could be attributed to a lingering mnemonic representation, decoupled from a motor one. Taken together, our results suggest that participants are predicting the specific visual content of an upcoming scene view—not just preparing an open/closed motor response—across head turns.

An interesting question for future study is the extent to which attentional shifts and visual predictions interact across large-scale visual actions. In experiment 3, the arrow cues preceding the prime image could have caused participants to prepare an attentional shift in working memory in advance of the head turn, consistent with past work showing that attentional shifts precede visual actions.[46–48] This likely attention shift aligns with—and adds to—the idea that visual predictions are prepared across head turns.

Across two separate studies, we provide evidence that memory-based predictions are generated for upcoming scene views across head turns. In these studies, presenting an incongruent prime view, as in the neutral and different-scene (experiments 1 and 3) conditions, creates a visual discontinuity between the prime and target image that does not exist in same-scene trials, where prime and target are contiguous views from the same scene. Such a discontinuity could slow response times, accounting for the response time difference between same-scene and discontinuous trials in experiments 1 and 3. However, our results argue against the impact of this potential: we found no difference in response time between the same-scene vs. neutral or different-scene conditions in experiment 2 (unfamiliar scenes), which also contain this visual discontinuity. Thus, the results of experiment 2 suggest that a violation of the predicted (remembered) scene view, rather than visual discontinuity between the prime and target images, likely underpins the priming effect observed in this study.

The current paradigm tackles a key challenge for memory-based predictions during active vision: the fact that perception and memory often operate in distinct spatial reference frames.[49] For example, consider a seated agent who places a cup of coffee on her desk—centering the cup on her retina—before turning to her left to work on a computer monitor. For the memory of the location of the coffee cup to be useful after the agent looks away, she needs to convert the coffee cup's position from retinotopic to spatiotopic coordinates—otherwise, she risks accidentally spilling the coffee while reaching for other items on the desk. Importantly, because memory-based predictions have primarily been studied in screen-based contexts,[2–4,8,9,50–52] it was previously unknown whether they are in fact implemented in a spatiotopic reference frame. Our findings show that memory-based predictions are mapped to the spatial coordinate frame of the
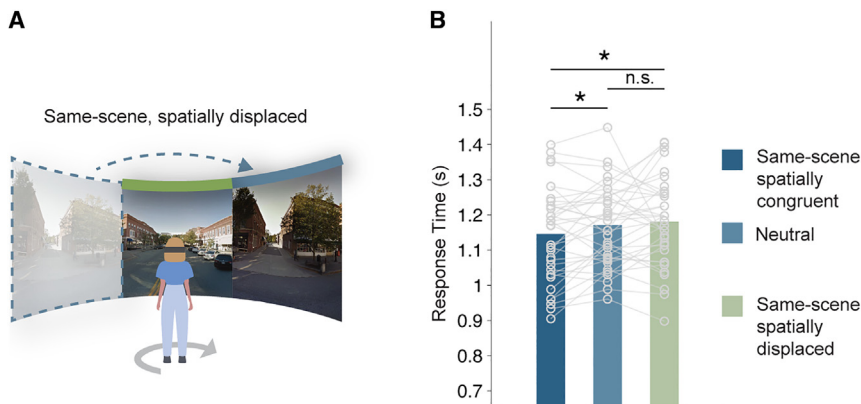
A

Same-scene, spatially displaced

B

Figure 4. Action dictates primed content

(A) Experiment 4 (*n* = 37) contained same-scene spatially displaced trials in which the prime and target were drawn from the same scene, but the target was presented 180° opposite its learned location in 360° space.

(B) Same-scene spatially displaced primes showed no priming effect relative to neutral. Connected points represent the same participant across conditions. * $p \leq 0.05$ difference between median response time for indicated priming conditions (paired t tests).

See also Tables S1 and S2 for additional analysis and Videos S1 and S2 for screen videos of training phase and priming test.

immediate global environment. This aligns with recent VR studies indicating that objects maintained in immersive working memory are primarily remembered based on their location in the environment's global coordinate frame (rather than on their egocentric position relative to the viewer during encoding).[49,53,54] A noteworthy question is whether memory-based visual predictions also map on to the broader, global coordinate frame when agents translate through their environment rather than simply turning in one discrete location as we have studied here. This question remains an intriguing one for future study. Together with our findings, VR studies underscore the role of memory in supporting ecologically relevant perceptual judgments[1,54] and highlight the promise of head-mounted VR as a tool for studying how perceptual information is represented—and predicted—during naturalistic, active perception.[34,55–57]

To conclude, here we developed a novel VR priming paradigm to investigate the role of memory-based predictions during naturalistic behavior. Our results indicate that memory-based predictions of upcoming scene views enable rapid perceptual judgments across head turns. We propose that memory-based predictions may be used to overcome our limited field of view by facilitating visual perception across large-scale visual actions (head turns and body movements). Future work should aim to understand how memory-based predictions retain their spatial structure, perhaps through interplay between the perceptual and head-direction systems.

RESOURCE AVAILABILITY

Lead contact
Requests for further information and resources should be directed to and will be fulfilled by the lead contact, Anna Mynick (annamynick@gmail.com).

Materials availability
This study did not generate new unique reagents.

Data and code availability
- Behavioral data have been deposited at Figshare and are publicly available as of the date of publication at https://doi.org/10.6084/m9.figshare.27655014.
- All original code has been deposited at GitHub and is publicly available at https://doi.org/10.6084/m9.figshare.27679719 as of the date of publication.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

AUTHOR CONTRIBUTIONS

C.E.R. and A.M. designed research; A.M. and T.L.B. developed software; A.M., A.J., and A.B. performed research; A.M. analyzed data; A.M., C.E.R., and A.S. wrote the paper.

DECLARATION OF INTERESTS

The authors declare no competing interests.

DECLARATION OF GENERATIVE AI AND AI-ASSISTED TECHNOLOGIES IN THE WRITING PROCESS

During the preparation of this work the authors used ChatGPT in order to improve the clarity of select sentences and phrases. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
- METHOD DETAILS
  - Hardware
  - Stimuli
  - Procedure
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Inclusion Criteria
  - Statistics

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cub.2024.11.024.

REFERENCES

1. Rust, N.C., and Palmer, S.E. (2021). Remembering the Past to See the Future. Annu. Rev. Vis. Sci. 7, 349–365. https://doi.org/10.1146/annurev-vision-093019-112249.

2. Hindy, N.C., Ng, F.Y., and Turk-Browne, N.B. (2016). Linking pattern completion in the hippocampus to predictive coding in visual cortex. Nat. Neurosci. 19, 665–667. https://doi.org/10.1038/nn.4284.

3. Kok, P., Rait, L.I., and Turk-Browne, N.B. (2020). Content-based Dissociation of Hippocampal Involvement in Prediction. J. Cogn. Neurosci. 32, 527–545. https://doi.org/10.1162/jocn_a_01509.

4. Kok, P., and Turk-Browne, N.B. (2018). Associative Prediction of Visual Shape in the Hippocampus. J. Neurosci. 38, 6888–6899. https://doi.org/10.1523/JNEUROSCI.0163-18.2018.

5. Turk-Browne, N.B., Scholl, B.J., Johnson, M.K., and Chun, M.M. (2010). Implicit Perceptual Anticipation Triggered by Statistical Learning. J. Neurosci. 30, 11177–11187. https://doi.org/10.1523/JNEUROSCI.0858-10.2010.

6. Olson, I.R., and Chun, M.M. (2001). Temporal contextual cuing of visual attention. J. Exp. Psychol. Learn. Mem. Cogn. 27, 1299–1313. https://doi.org/10.1037/0278-7393.27.5.1299.

7. Nissen, M.J., and Bullemer, P. (1987). Attentional requirements of learning: evidence from performance measures. Cogn. Psychol. 19, 1–32. https://doi.org/10.1016/0010-0285(87)90002-8.

8. Cravo, A.M., Rohenkohl, G., Santos, K.M., and Nobre, A.C. (2017). Temporal Anticipation Based on Memory. J. Cogn. Neurosci. 29, 2081–2089.

9. Stokes, M.G., Atherton, K., Patai, E.Z., and Nobre, A.C. (2012). Long-term memory prepares neural activity for perception. Proc. Natl. Acad. Sci. USA 109, E360–E367. https://doi.org/10.1073/pnas.1108555108.

10. Chun, M.M., and Jiang, Y. (1998). Contextual Cueing: Implicit Learning and Memory of Visual Context Guides Spatial Attention. Cogn. Psychol. 36, 28–71. https://doi.org/10.1006/cogp.1998.0681.

11. Jaramillo, S., and Zador, A. (2010). Auditory cortex mediates the perceptual effects of acoustic temporal expectation. Nat. Neurosci. 14, 246–251. https://doi.org/10.1038/npre.2010.5139.1.

12. Kroell, L.M., and Rolfs, M. (2022). Foveal vision anticipates defining features of eye movement targets. eLife 11, e78106. https://doi.org/10.7554/eLife.78106.

13. Herwig, A., and Schneider, W.X. (2014). Predicting object features across saccades: Evidence from object recognition and visual search. J. Exp. Psychol. Gen. 143, 1903–1922. https://doi.org/10.1037/a0036781.

14. Osterbrink, C., and Herwig, A. (2021). Prediction of complex stimuli across saccades. J. Vis. 21, 10. https://doi.org/10.1167/jov.21.2.10.

15. Duhamel, J.R., Colby, C.L., and Goldberg, M.E. (1992). The Updating of the Representation of Visual Space in Parietal Cortex by Intended Eye Movements. Science 255, 90–92. https://doi.org/10.1126/science.1553535.

16. Umeno, M.M., and Goldberg, M.E. (1997). Spatial Processing in the Monkey Frontal Eye Field. I. Predictive Visual Responses. J. Neurophysiol. 78, 1373–1383. https://doi.org/10.1152/jn.1997.78.3.1373.

17. Walker, M.F., Fitzgibbon, E.J., and Goldberg, M.E. (1995). Neurons in the monkey superior colliculus predict the visual result of impending saccadic eye movements. J. Neurophysiol. 73, 1988–2003. https://doi.org/10.1152/jn.1995.73.5.1988.

18. Denagamage, S., Morton, M.P., and Nandy, A.S. (2024). Widespread Receptive Field Remapping in Early Visual Cortex. Preprint at bioRxiv.

19. Merriam, E.P., Genovese, C.R., and Colby, C.L. (2007). Remapping in Human Visual Cortex. J. Neurophysiol. 97, 1738–1755. https://doi.org/10.1152/jn.00189.2006.

20. Parks, N.A., and Corballis, P.M. (2010). Human transsaccadic visual processing: Presaccadic remapping and postsaccadic updating. Neuropsychologia 48, 3451–3458. https://doi.org/10.1016/j.neuropsychologia.2010.07.028.

21. Parks, N.A., and Corballis, P.M. (2008). Electrophysiological correlates of presaccadic remapping in humans. Psychophysiology 45, 776–783. https://doi.org/10.1111/j.1469-8986.2008.00669.x.

22. He, D., Mo, C., and Fang, F. (2017). Predictive feature remapping before saccadic eye movements. J. Vis. 17, 14. https://doi.org/10.1167/17.5.14.

23. He, T., Fritsche, M., and De Lange, F.P. (2018). Predictive remapping of visual features beyond saccadic targets. J. Vis. 18, 20. https://doi.org/10.1167/18.13.20.

24. Melcher, D. (2007). Predictive remapping of visual features precedes saccadic eye movements. Nat. Neurosci. 10, 903–907. https://doi.org/10.1038/nn1917.

25. Rolfs, M., Jonikaitis, D., Deubel, H., and Cavanagh, P. (2011). Predictive remapping of attention across eye movements. Nat. Neurosci. 14, 252–256. https://doi.org/10.1038/nn.2711.

26. Chadwick, M.J., Mullally, S.L., and Maguire, E.A. (2013). The hippocampus extrapolates beyond the view in scenes: An fMRI study of boundary extension. Cortex 49, 2067–2079. https://doi.org/10.1016/j.cortex.2012.11.010.

27. Intraub, H., and Richardson, M. (1989). Wide-Angle Memories of Close-Up Scenes. J. Exp. Psychol. Learn. Mem. Cogn. 15, 179–187.

28. Mullally, S.L., Intraub, H., and Maguire, E.A. (2012). Attenuated Boundary Extension Produces a Paradoxical Memory Advantage in Amnesic Patients. Curr. Biol. 22, 261–268. https://doi.org/10.1016/j.cub.2012.01.001.

29. Park, S., Intraub, H., Yi, D.-J., Widders, D., and Chun, M.M. (2007). Beyond the Edges of a View: Boundary Extension in Human Scene-Selective Visual Cortex. Neuron 54, 335–342. https://doi.org/10.1016/j.neuron.2007.04.006.

30. Bainbridge, W.A., and Baker, C.I. (2020). Boundaries Extend and Contract in Scene Memory Depending on Image Properties. Curr. Biol. 30, 537–543.e3. https://doi.org/10.1016/j.cub.2019.12.004.

31. Gandolfo, M., Nägele, H., and Peelen, M.V. (2023). Predictive Processing of Scene Layout Depends on Naturalistic Depth of Field. Psychol. Sci. 34, 394–405.

32. Hafri, A., Wadhwa, S., and Bonner, M.F. (2022). Perceived Distance Alters Memory for Scene Boundaries. Psychol. Sci. 33, 2040–2058.

33. Berens, S.C., Joensen, B.H., and Horner, A.J. (2021). Tracking the Emergence of Location-based Spatial Representations in Human Scene-Selective Cortex. J. Cogn. Neurosci. 33, 445–462. https://doi.org/10.1162/jocn_a_01654.

34. Robertson, C.E., Hermann, K.L., Mynick, A., Kravitz, D.J., and Kanwisher, N. (2016). Neural Representations Integrate the Current Field of View with the Remembered 360° Panorama in Scene-Selective Cortex. Curr. Biol. 26, 2463–2468. https://doi.org/10.1016/j.cub.2016.07.002.

35. Matthis, J.S., Yates, J.L., and Hayhoe, M.M. (2018). Gaze and the Control of Foot Placement When Walking in Natural Terrain. Curr. Biol. 28, 1224–1233.e5. https://doi.org/10.1016/j.cub.2018.03.008.

36. Hayhoe, M.M. (2017). Vision and Action. Annu Rev Vis Sci 3, 389–413. https://doi.org/10.1146/annurev-vision-102016-061437.

37. Castelhano, M.S., and Pollatsek, A. (2010). Extrapolating spatial layout in scene representations. Mem. Cognit. 38, 1018–1025. https://doi.org/10.3758/MC.38.8.1018.

38. Sanocki, T. (2003). Representation and perception of scenic layout. Cogn. Psychol. 47, 43–86. https://doi.org/10.1016/S0010-0285(03)00002-1.

39. Sanocki, T., and Epstein, W. (1997). Priming spatial layout of scenes. Psychol. Sci. 8, 374–378.

40. Shafer-Skelton, A., and Brady, T.F. (2019). Scene layout priming relies primarily on low-level features rather than scene layout. J. Vis. 19, 14. https://doi.org/10.1167/19.1.14.

41. Nobre, A.C., and Stokes, M.G. (2019). Premembering Experience: A Hierarchy of Time-Scales for Proactive Attention. Neuron *104*, 132–146. https://doi.org/10.1016/j.neuron.2019.08.030.

42. Rappaport, S.J., Humphreys, G.W., and Riddoch, M.J. (2013). The attraction of yellow corn: reduced attentional constraints on coding learned conjunctive relations. J. Exp. Psychol. Hum. Percept. Perform. *39*, 1016–1031. https://doi.org/10.1037/a0032506.

43. Albright, T.D. (2012). On the Perception of Probable Things: Neural Substrates of Associative Memory, Imagery, and Perception. Neuron *74*, 227–245. https://doi.org/10.1016/j.neuron.2012.04.001.

44. Schlack, A., and Albright, T.D. (2007). Remembering Visual Motion: Neural Correlates of Associative Plasticity and Motion Recall in Cortical Area MT. Neuron *53*, 881–890. https://doi.org/10.1016/j.neuron.2007.02.028.

45. de Lange, F.P., Heilbron, M., and Kok, P. (2018). How Do Expectations Shape Perception? Trends Cogn. Sci. *22*, 764–779. https://doi.org/10.1016/j.tics.2018.06.002.

46. Deubel, H. (2008). The time course of presaccadic attention shifts. Psychol. Res. *72*, 630–640. https://doi.org/10.1007/s00426-008-0165-3.

47. Gresch, D., Boettcher, S.E.P., Van Ede, F., and Nobre, A.C. (2024). Shifting attention between perception and working memory. Cognition *245*, 105731. https://doi.org/10.1016/j.cognition.2024.105731.

48. Harrison, W.J., Mattingley, J.B., and Remington, R.W. (2012). Pre-Saccadic Shifts of Visual Attention. PLoS One *7*, e45670. https://doi.org/10.1371/journal.pone.0045670.

49. Draschkow, D., Nobre, A.C., and van Ede, F. (2022). Multiple spatial frames for immersive working memory. Nat. Hum. Behav. *6*, 536–544. https://doi.org/10.1038/s41562-021-01245-y.

50. Aitken, F., and Kok, P. (2022). Hippocampal representations switch from errors to predictions during acquisition of predictive associations. Nat. Commun. *13*, 3294. https://doi.org/10.1038/s41467-022-31040-w.

51. Clarke, A., Crivelli-Decker, J., and Ranganath, C. (2022). Contextual expectations shape cortical reinstatement of sensory representations. J. Neurosci. *42*, 5956–5965. https://doi.org/10.1523/JNEUROSCI.2045-21.2022.

52. Crivelli-Decker, J., Clarke, A., Park, S.A., Huffman, D.J., Boorman, E., and Ranganath, C. (2021). Goal-centered representations in the human hippocampus. Nat. Commun. *14*, 2946.

53. Thom, J.L., Nobre, A.C., Van Ede, F., and Draschkow, D. (2023). Heading Direction Tracks Internally Directed Selective Attention in Visual Working Memory. J. Cogn. Neurosci. *35*, 856–868. https://doi.org/10.1162/jocn_a_01976.

54. Van Ede, F., and Nobre, A.C. (2023). Turning Attention Inside Out: How Working Memory Serves Behavior. Annu. Rev. Psychol. *74*, 137–165. https://doi.org/10.1146/annurev-psych-021422-041757.

55. Botch, T.L., Garcia, B.D., Choi, Y.B., Feffer, N., and Robertson, C.E. (2023). Active visual search in naturalistic environments reflects individual differences in classic visual search performance. Sci. Rep. *13*, 631. https://doi.org/10.1038/s41598-023-27896-7.

56. Cohen, M.A., Botch, T.L., and Robertson, C.E. (2020). The limits of color awareness during active, real-world vision. Proc. Natl. Acad. Sci. USA *117*, 13821–13827. https://doi.org/10.1073/pnas.1922294117.

57. Haskins, A.J., Mentch, J., Botch, T.L., and Robertson, C.E. (2020). Active vision in immersive, 360° real-world environments. Sci. Rep. *10*, 14304. https://doi.org/10.1038/s41598-020-71125-4.

58. Hansen, N.E., Noesen, B.T., Nador, J.D., and Harel, A. (2018). The influence of behavioral relevance on the processing of global scene properties: An ERP study. Neuropsychologia *114*, 168–180. https://doi.org/10.1016/j.neuropsychologia.2018.04.040.

59. Harel, A., Kravitz, D.J., and Baker, C.I. (2013). Deconstructing Visual Scenes in Cortex: Gradients of Object and Spatial Layout Information. Cereb. Cortex *23*, 947–957. https://doi.org/10.1093/cercor/bhs091.

60. Harel, A., Groen, I.I.A., Kravitz, D.J., Deouell, L.Y., and Baker, C.I. (2016). The Temporal Dynamics of Scene Processing: A Multifaceted EEG Investigation. eNeuro *3*, ENEURO.0139-16.2016. https://doi.org/10.1523/ENEURO.0139-16.2016.

61. Kravitz, D.J., Peng, C.S., and Baker, C.I. (2011). Real-World Scene Representations in High-Level Visual Cortex: It's the Spaces More Than the Places. J. Neurosci. *31*, 7322–7333. https://doi.org/10.1523/JNEUROSCI.4588-10.2011.

62. Greene, M.R., and Oliva, A. (2009). Recognition of natural scenes from global properties: Seeing the forest without representing the trees. Cogn. Psychol. *58*, 137–176. https://doi.org/10.1016/j.cogpsych.2008.06.001.

63. Kuznetsova, A., Brockhoff, P.B., and Christensen, R.H.B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. J. Stat. Softw. *82*, 1–26. https://doi.org/10.18637/jss.v082.i13.

64. Balota, D.A., Aschenbrenner, A.J., and Yap, M.J. (2013). Additive effects of word frequency and stimulus quality: The influence of trial history and data transformations. J. Exp. Psychol. Learn. Mem. Cogn. *39*, 1563–1571. https://doi.org/10.1037/a0032186.

65. Masson, M.E.J., and Kliegl, R. (2013). Modulation of additive and interactive effects in lexical decision by trial history. J. Exp. Psychol. Learn. Mem. Cogn. *39*, 898–914. https://doi.org/10.1037/a0029180.

66. Lüdecke, D., Ben-Shachar, M.S., Patil, I., Waggoner, P., and Makowski, D. (2021). performance: An R package for assessment, comparison and testing of statistical models. J. Open Source Softw. *6*, 3139.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Raw data	This paper; https://doi.org/10.6084/m9.figshare.27655014	https://doi.org/10.6084/m9.figshare.27655014
Software and algorithms		
MATLAB 2024a	MathWorks	https://www.mathworks.com/products/matlab.html
RStudio 2024.09.0+375	Posit	https://posit.co/download/rstudio-desktop/
R Statistical Software	R Project	www.r-project.org
Contributed R packages	Comprehensive R Archive Network (CRAN)	cran.r-project.org
Unity 2018.4.12f1	Unity	https://unity.com
ManageXR	ManageXR	https://www.managexr.com
Custom-built MATLAB and R code	This paper; https://doi.org/10.6084/m9.figshare.27679719	https://doi.org/10.6084/m9.figshare.27679719
Other		
Oculus Quest 2	Meta	https://www.meta.com/quest/products/quest-2/

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

140 participants were recruited across four experiments in this study. Sample size for experiments 1-2 was chosen based on previous work[34]. For experiments 3-4, a power analysis based on Pilot B (see Figure S4) was conducted (power = 0.8, alpha = 0.05, effect size = 0.56, paired, one-tailed) to determine the sample size. Adult participants were recruited from the local university or the local community and received payment ($15/hour) or course credit for participation. Due to technical challenges of remote data collection, age and sex were not recorded for experiment 1, however these participants indicated they were between 18-65 years old during consent. Experiment 2 (mean age: 22.84 ± 5.09 years) had 15 female and 4 male participants, and 1 participant who responded, "Prefer not to say". Experiment 3 (mean age: 20.23 ± 3.94 years) had 16 female and 10 male participants. Experiment 4 (mean age: 20.54 ± 2.82 years) had 22 female and 15 male participants. Participants in experiments 2-4 were recruited based on the following criteria: 1) normal or corrected-to-normal vision 2) no neurological or psychiatric conditions 3) no history of epilepsy. Participants in experiment 1 participated for course credit and completed the experiment remotely from their homes using virtual reality headsets that were delivered to them in the mail. Participants in experiments 2 and 3 were mixed between remote testing and in-person testing at the local university in the following amounts: experiment 2: 7 remote, 17 in-person; experiment 3: 4 remote, 42 in-person. Participants in experiment 4 all participated in person. Remote data was collected to comply with health and safety guidelines during the COVID-19 pandemic. There was no significant impact of remote versus in-person testing in any of the experiments (Table S3). Written consent was obtained from all participants in accordance with a protocol approved by the local university Institutional Review Board. 32 participants were excluded due to data quality concerns associated with VR data collection (see inclusion criteria section below), leaving data from 101 participants in the final analyses: 18 in experiment 1, 20 in experiment 2, 26 in experiment 3, and 37 in experiment 4.

METHOD DETAILS

Hardware

This study took place entirely in head-mounted virtual reality (VR). Stimuli were displayed through head-mounted virtual reality (Oculus Quest 2, resolution 1832x1920 per eye, approximately 97° horizontal by 93° vertical field of view, 120 Hz refresh rate) using a project designed in Unity (www.unity3d.com) with custom scripts written in C#, and deployed to headsets using mobile device management software ManageXR (www.managexr.com). Experimental data was transferred from the HMD to lab servers via custom data transfer pipeline written in C# and PHP.[55]

Stimuli

Stimuli were real-world photospheres downloaded from Google Maps. The number of indoor/outdoor locations was balanced in each experiment. In all experiments, scenes had a specific layout: each contained an open side (e.g. a field or parking lot) and a closed side (e.g. a wall or facade) on opposite sides of an initial facing direction (i.e. separated by 180°) (Figure S1). In all experiments, we counterbalanced the number of [open-left, closed-right] and [closed-left, open-right] scenes.

Procedure

Our study consisted of 4 experiments that each tested a specific question related to memory-based predictions in immersive environments. All experiments used a novel, shared priming test inspired by past priming tests of associative memory[34] to evaluate the presence or absence of prediction-based facilitation of an ecologically-relevant perceptual judgment.

This specific open/closed task was chosen because (i) scene expanse judgments (open/closed) are commonly used in the scene literature[58–60] and are thought to be a key representational dimension in scene-selective cortex,[61] (ii) because open/closed judgments required participants to allocate their attention broadly to the target scene image, as opposed to searching for a specific object or detail,[62] and (iii) because it was important that prime view alone could not predict the correct answer for the target image, which ruled out other common scene properties (indoor/outdoor, scene category, manmade/natural), which tend to hold constant across all views of a scene.

In the sections that follow, we describe the procedure for each experiment.

Experiment 1 Procedure Overview

Experiment 1 was conducted over two consecutive days. On day 1, participants studied the scenes in detail (training phase). Then, they received instructions on distinguishing between open and closed scene views (Open/Closed Instructions). On day 2, participants performed the priming test.

Experiment 1 Training Phase

During the training phase on day 1, participants actively viewed 18 immersive photorealistic scenes taken from the area around the local university's campus for 20 seconds each (see Figure S1). The open/closed sides of a scene were always positioned to participants' left/right relative to their starting facing direction and the back-most 90° of each scene was occluded to provide a clear sense of left/center/right in each scene. The occluder was necessary to provide a framework to subsequently probe participants' memory for the spatial layout of each scene (see Experiment 1 Explicit Memory Test). Participants were asked to "complete this study standing at a desk or countertop where you can comfortably reach your keypad." For each scene, participants were instructed to "Look around like you normally would". The instructions told participants that there would be a "gray wall" (occluder) behind them in each scene, and that they only needed to explore what wasn't blocked by the wall. After viewing each scene, they indicated whether the scene was familiar in real life via wireless keypad (1=Yes, 2=No, 3=Unsure) (Video S3). One participant in experiment 1 was not included in the calculation of percent familiarity (Figure S2) because of missing data due to technical challenges with remote data collection. On day 2, participants completed a refresher training phase in which they viewed the 18 scenes again for 20 seconds each.

Experiment 1 Open/Closed Instructions

Our primary interest was participants' ability to make an ecologically-relevant behavioral judgment: was a scene view navigable or not. To this end, after viewing each scene on day 1, we instructed participants how to distinguish between navigable vs. non-navigable (i.e., open vs. closed) scene views. Before receiving these instructions, participants were informed that open/closed judgments would be performed in a subsequent part of the experiment. Our specific instructions for distinguishing open versus closed scene views were: "If a snapshot shows an open space or pathway where you could walk into the scene and move forward freely, that snapshot is Open. If a snapshot shows something blocking your way – something that would prevent you from walking into the scene and moving forward freely – that snapshot is Closed." (Video S2). We then showed participants several examples of open versus closed scene views in novel, unstudied scenes. After viewing these examples, participants completed 6 open/closed judgment practice trials with feedback.

On day 2, following the refresher training phase, participants were reminded of the instructions for distinguishing open versus closed scene views using the same wording as on day 1.

Experiment 1 Priming Test

On day 2, participants completed the priming test after the refresher Training and the Open/Closed instructions reminder. Each trial of the priming test consisted of four events (Figure 1): 1) Prime image (110° wide, 180° tall) appears directly in front of the participant (display time: 300 ms); 2) Target image (110° wide, 180° tall) appears to the left or right of the participant (visible in their periphery); 3) Participant turns 90° to face the target image; 4) Participant indicates whether the target is open or closed via button press on a wireless keypad.

The key manipulation of experiment 1 was controlling whether prime image came from the same scene as the target image. To encourage participants to use the prime image to predict the upcoming target, more trials contained same-scene primes (50% of 144 total trials) than neutral (25%) or different-scene (25%) primes. The neutral prime for this study was simply a grey screen. The 36 possible target views (drawn from 2 sides of 18 scenes) occurred twice each in the same-scene condition and once each in the neutral condition. In the different-scene condition, the prime was drawn from each of the 18 possible scene center views two times; the target was drawn from each of the 36 possible target views one time. The mapping between prime and target views was randomized during trial order creation, but trial order was constant across participants.

Before testing, participants completed 5 practice priming test trials with feedback on novel, unstudied scenes. We instructed participants to respond as quickly as they knew the answer using a wireless keypad (4=Open, 3=Closed) operated using two fingers on one hand, and to "make sure your fingers are ready on the keypad". After each trial, the participants' response time was displayed to emphasize the importance of responding quickly. Because the open/closed task was relatively simple, and participants were highly trained to perform it, this study focuses on differences in response time rather than accuracy across conditions.

Experiment 1 Explicit Memory Test

After the priming test, we tested participants' explicit memory of each scene's spatial structure. On each trial of the explicit memory test, participants saw a 45° view from a studied scene directly in front of them. Participants then used a wireless keypad to indicate whether that view appeared on the Left, Right, or Center of the scene during the training phase (Figure S3). Participants were tested on each left/right/center view for each studied scene.

Experiment 2 Procedure Overview

Experiment 2 specifically investigated whether priming occurred in unfamiliar environments. For this reason, unlike experiment 1, experiment 2 required no training phase, and took place on a single day. In experiment 2, participants first learned Open/Closed instructions, and then performed the priming test on unfamiliar scenes.

Experiment 2 Open/Closed Instructions

We instructed participants how to distinguish between navigable vs. non-navigable (i.e., open vs. closed) scene views in the same way as in experiment 1. Before receiving these instructions, participants were informed that open/closed judgments would be performed in a subsequent part of the experiment. As in all other experiments, our specific instructions for distinguishing open versus closed scene views were: "If a snapshot shows an open space or pathway where you could walk into the scene and move forward freely, that snapshot is Open. If a snapshot shows something blocking your way – something that would prevent you from walking into the scene and moving forward freely – that snapshot is Closed." (Video S2). We then showed participants several examples of open versus closed scene views in novel, unstudied scenes. After viewing these examples, participants completed 6 open/closed judgment practice trials with feedback.

Experiment 2 Priming Test

After receiving Open/Closed instructions, participants completed the priming test. Importantly, unlike experiments 1 and 3, this test took place on novel (unfamiliar) scenes without any training phase. Identical to experiment 1, each trial of the priming test consisted of four events (Figure 1): 1) Prime image (110° wide, 180° tall) appeared directly in front of the participant (display time: 300 ms); 2) Target image (110° wide, 180° tall) appeared to the left or right of the participant (visible in their periphery); 3) Participant turned 90° to face the target image; 4) Participant indicated whether the target was open or closed via button press on a wireless keypad.

In experiment 2, the key manipulation was using unfamiliar scenes for which participants had no memory. Thus, experiment 2 prime and target images were drawn from diverse locations around the world (Figure S5). To match the condition ratios of experiment 1, prime images were: same-scene primes (50% of 144 total trials), neutral (25%) or different-scene (25%) primes. The neutral prime for this study was simply a grey screen. The 72 possible target views (drawn from 2 sides of 36 scenes) occurred once each in the same-scene condition and were randomly split between assignment to either the neutral or different-scene conditions. The mapping between prime and target views was randomized during trial order creation, but trial order was constant across participants.

Before testing, participants completed 4 practice priming test trials with feedback on novel scenes independent from the test set. As in all other experiments, we instructed participants to respond as quickly as they knew the answer using a wireless keypad (4=Open, 3=Closed) operated using two fingers on one hand, and to "make sure your fingers are ready on the keypad". After each trial, the participants' response time was displayed to emphasize the importance of responding quickly.

Experiment 3 Procedure Overview

Like experiment 1, experiment 3 was conducted over two consecutive days. On day 1, participants studied the scenes in detail (training phase). Then, they received instructions on distinguishing between open and closed scene views (Open/Closed Instructions). On day 2, participants performed the priming test.

Experiment 3 Training Phase

As in experiment 1, during the experiment 3 training phase on day 1, participants actively viewed 18 immersive photorealistic scenes taken from the area around the local university's campus for 20 seconds each (see Figure S1). As in all experiments, the open/closed sides of a scene were always positioned to participants' left/right relative to their starting facing direction (presentation side counterbalanced), and the back-most 90° of each scene was occluded to provide a clear sense of left/center/right in each scene. Consistent with the training phase in experiment 1, participants were asked to "complete this study standing at a desk or countertop where you can comfortably reach your keypad." For each scene, participants were instructed to "Look around like you normally would". The instructions told participants that there would be a "gray wall" (occluder) behind them in each scene, and that they only needed to explore what wasn't blocked by the wall. After viewing each scene, they indicated whether the scene was familiar in real life via wireless keypad (1=Yes, 2=No, 3=Unsure). Three participants in experiment 3 were not included in the calculation of percent familiarity (Figure S2) because of missing data due to technical challenges with remote data collection.

On day 2, participants completed a refresher training phase in which they viewed the 18 scenes again for 20 seconds each.

Experiment 3 Open/Closed Instructions

As in experiment 1, on day 1, we instructed participants how to distinguish between navigable vs. non-navigable (i.e., open vs. closed) scene views. Before receiving these instructions, participants were informed that open/closed judgments would be performed in a subsequent part of the experiment. Our specific instructions for distinguishing open versus closed scene views were: "If a snapshot shows an open space or pathway where you could walk into the scene and move forward freely, that snapshot is Open. If a snapshot shows something blocking your way – something that would prevent you from walking into the scene and moving forward freely – that snapshot is Closed." (Video S2). We then showed participants several examples of open versus closed scene views in novel, unstudied scenes. After viewing these examples, participants completed 6 open/closed judgment practice trials with feedback.

Participants were reminded of the instructions for distinguishing open versus closed scene views on day 2, following the refresher training phase, using the same wording as on day 1.

Experiment 3 Direction-cued Priming Test

Experiment 3 tested whether priming was skewed in the direction of intended head turns. For this reason, we modified the standard priming test (used in experiments 1 and 2) to manipulate the anticipated direction needed to turn and face the target image.

The trial sequence in experiment 3 consisted of 5 stages (Figure 3): 1) Participants saw an arrow indicating the direction to plan a head turn in to see the target image (300 ms display time). On 2/3 of trials, the arrow correctly indicated the direction of the upcoming target (i.e., was a valid cue), and 1/3 of trials the arrow indicated the incorrect direction (i.e., was invalid. 2). Participants then saw the prime image (300 ms display time). 3) The target image (110° wide, 180° tall) appeared to the participants' left or right. 4) Participants turned 90° to face the target. 5) Participants responded using a button press.

Prime images in experiment 3 were either same-scene or neutral, resulting in four priming conditions: same-scene valid arrow (50% of 216 trials), neutral valid arrow (16.7%), same-scene invalid arrow (16.7%) and neutral invalid arrow (16.7%). Neutral prime images were scrambled scenes (32 x 16 grid). The scrambled neutral prime was introduced to better match the amount of visual information across priming conditions. The 36 possible target views (drawn from 2 sides of 18 scenes) occurred three times each in the same-scene, valid arrow condition and once each in (i) same-scene, invalid arrow, (ii) different-scene, valid arrow and (iii) different-scene, invalid arrow conditions. The mapping between prime and target views was randomized during trial order creation, but trial order was constant across participants.

Before testing, participants completed 6 practice priming test trials with feedback on novel, unstudied scenes.

Experiment 4 Procedure Overview

Like experiments 1 and 3, experiment 4 was conducted over two consecutive days. The key manipulation in experiment 4 was introducing the "same-scene spatially displaced" condition to assess whether predicted content depended on participants' actions within the environment. For clarity, we called the same-scene condition (experiments 1, 2 and 3) "same-scene spatially congruent" in experiment 4. On day 1, participants studied the scenes in detail (training phase). Then, they received instructions on distinguishing between open and closed scene views (Open/Closed Instructions). On day 2, participants performed the priming test.

Experiment 4 Training Phase

As in experiments 1 and 3, during the experiment 4 training phase on day 1, participants actively viewed 18 immersive photorealistic scenes taken from the area around the local university's campus for 20 seconds each (see Figure S1). As in all experiments, the open/closed sides of a scene were always positioned to participants' left/right relative to their starting facing direction and the back-most 90° of each scene was occluded to provide a clear sense of left/center/right in each scene. Participants were asked to "complete this study standing at a desk or countertop where you can comfortably reach your keypad." For each scene, participants were instructed to "Look around like you normally would". The instructions told participants that there would be a "gray wall" (occluder) behind them in each scene, and that they only needed to explore what wasn't blocked by the wall. After viewing each scene, they indicated whether the scene was familiar in real life via wireless keypad (1=Yes, 2=No, 3=Unsure). On day 2, participants completed a refresher training phase in which they viewed the 18 scenes again for 20 seconds each.

Experiment 4 Open/Closed Instructions

As in experiment 1, on day 1, we instructed participants how to distinguish between navigable vs. non-navigable (i.e., open vs. closed) scene views. Before receiving these instructions, participants were informed that open/closed judgments would be performed in a subsequent part of the experiment. As in all other experiments, our specific instructions for distinguishing open versus closed scene views were: "If a snapshot shows an open space or pathway where you could walk into the scene and move forward freely, that snapshot is Open. If a snapshot shows something blocking your way – something that would prevent you from walking into the scene and moving forward freely – that snapshot is Closed." (Video S2). We then showed participants several examples of open versus closed scene views in novel, unstudied scenes. After viewing these examples, participants completed 6 open/closed judgment practice trials with feedback.

Experiment 4 Priming Test

On day 2, participants completed the priming test after the refresher Training and the Open/Closed instructions reminder. Each trial of the priming test consisted of four events (Figure 1): 1) Prime image (110° wide, 180° tall) appeared directly in front of the participant (display time: 300 ms); 2) Target image (110° wide, 180° tall) appeared to the left or right of the participant (visible in their periphery); 3) Participant turned 90° to face the target image; 4) Participant indicated whether the target was open or closed via button press on a wireless keypad.

In addition to same-scene and neutral conditions, experiment 4 introduced the same-scene spatially displaced priming condition, in which targets were drawn from the same scene as the prime, but were presented 180° their learned spatial location. As in experiment 1, to encourage participants to use the prime image to predict the upcoming target, more trials contained same-scene primes (50% of 144 total trials) than neutral (25%) or same-scene spatially displaced (25%) primes. The neutral prime for this study was a grey screen. The 36 possible target views (drawn from 2 sides of 18 scenes) occurred twice each in the same-scene condition and once each in the neutral condition. The mapping between prime and target views was randomized during trial order creation, but trial order was constant across participants.

Before testing, participants completed 5 practice priming test trials with feedback on novel, unstudied scenes. We instructed participants to respond as quickly as they knew the answer using a wireless keypad (4=Open, 3=Closed) operated using two fingers on one hand, and to "make sure your fingers are ready on the keypad". After each trial, the participants' response time was displayed to emphasize the importance of responding quickly.

Experiment 4 Explicit Memory Test

As in experiment 1, after the priming test, we tested participants' explicit memory of each scene's spatial structure. On each trial of the explicit memory test, participants saw a 45° view from a studied scene directly in front of them. Participants then used a wireless keypad to indicate whether that view appeared on the Left, Right, or Center of the scene during the training phase (Figure S3). Participants were tested on each left/right/center view for each studied scene.

QUANTIFICATION AND STATISTICAL ANALYSIS

We analyzed data using custom MATLAB scripts (www.mathworks.com) and custom R scripts.

Inclusion Criteria

Due to the novelty of the VR testing environment for remote testing, we applied a stringent trial-level inclusion criteria to ensure high data quality and to ensure that data from experiments could be equated. For instance, in all experiments, participants were required to view the prime for its entire duration (300ms) before turning to the target. However, unlike experiments 1, 2, and 4, the arrow cue in experiment 3 enabled participants to anticipate the upcoming location of the target, which led to more trial-level exclusions caused by participants turning before the end of the prime (despite instructions, as in all experiments, to wait for target onset). This tradeoff – stringent inclusion criteria over more tailored task instructions in experiment 3 – was critical for ensuring that participants received comparable instructions across experiments.

After applying these criteria, data were included in the following percentages per included participant: experiment 1: M = 74.92%, SD = 11.90%; experiment 2: 75.03%, SD = 7.07%; experiment 3: M = 79.20%, SD = 7.01%; experiment 4: M = 72.43%, SD = 8.56%.

Response time data from a priming test trial was excluded from analysis if the open/closed response was:

1. incorrect;
2. faster than 250ms, potentially indicating insufficient time spent to perform the judgment accurately;
3. outside of the participants' mean response time ± 3 standard deviations.

Trials were also excluded if:

1. the headset failed to transfer head-orientation data;
2. the participant turned before at least 50% of the prime image was within view;
3. the participant did not turn far enough toward the target to view at least 50% of the target scene segment;
4. the participant turned the wrong way (i.e. away from the target image) by more than 15 degrees following the prime (experiments 1,2 and 4).

In experiment 3, in trials where participants turned the wrong way (i.e. away from the target image) by more than 15 degrees:

1. If the participant corrected their head turn to the correct direction, the timepoint at which the participant corrected their head turn served as the start of the corrected trial.
2. If the timepoint at which the participant corrected their head turn could not be identified, the trial was excluded.

Finally, for a given participant, if more than 50% of trials in any priming condition were excluded or if their median response time exceeded ± 2 standard deviations from the group median, we removed that participant from data analysis entirely.

Statistics

For all experiments, we calculated the median response time for each priming test condition. Response times were calculated from the appearance of the target and included the time taken to turn to face the target.

For each priming test, we used linear mixed effects models to evaluate the effect of priming condition on response time using the 'lmerTest'[63] package in R. This approach was selected to reduce the opportunity for just a subset of scenes to account for the priming effect. In each model, we included priming condition as a fixed effect and participant and target scene identity as random effects

(intercepts). Raw response times were transformed[64,65] (inverse square root) to ensure normality of residuals and homogeneity of variance, confirmed using 'check model' in the 'performance' package in R.[66] Raw (i.e. untransformed) response times yielded the same pattern of results as the transformed data in all LMEs (Table S1). Likewise, pairwise t tests run on median transformed response times yielded the same pattern of results as median raw response times (Table S2). To evaluate the extent of priming in experiment 1 versus experiment 2, we added experiment as a fixed effect to our model. Where appropriate, we used two-tailed, pairwise Bonferroni-corrected t tests for post-hoc tests.