How sturdy is your memory palace? Reliable room representations predict subsequent reinstatement of placed objects

Rolando Masís-Obando[1,4*], Kenneth A. Norman[1,2] and Christopher Baldassano[3]

[1*]Princeton Neuroscience Institute, Princeton University, Princeton, NJ, USA.
[2]Department of Psychology, Princeton University, Princeton, NJ, USA.
[3]Department of Psychology, Columbia University, New York, NY, USA.
[4]Psychological & Brain Sciences, Johns Hopkins University, Baltimore, MD, USA.

*Corresponding author(s). E-mail(s): rmasis@princeton.edu;

Abstract

What are the neural properties that make spatial contexts effective scaffolds for storing and accessing memories? We hypothesized that spatial locations with stable and distinctive (i.e., reliable) neural representations would best support memory for new experiences. To test this, participants learned the layout of a custom-built 23-room virtual reality (VR) "memory palace" that they explored using a head-mounted display. The next day, participants underwent whole-brain fMRI while watching videos of the rooms, allowing us to measure the reliability of the neural activity pattern associated with each room. Participants then returned to VR to encode 23 objects placed in each of the 23 rooms and later recalled the rooms and objects during fMRI. We found that our room reliability measure (computed *prior to* encoding) predicted object reinstatement during recall across cortex; this was driven not only by group-level reliability across participants, but also idiosyncratic reliability within participants. Moreover, this effect did not arise through enhanced retrieval of reliable rooms during recall, since the relationship between reliability and object reinstatement remained significant when controlling for room reinstatement during retrieval; this suggests that, instead, room reliability promotes improved binding of rooms to objects at

encoding. Together, these results showcase how the quality of the neural representation of a spatial context can be quantified and used to 'audit' its utility as a memory scaffold for future experiences.

Keywords: spatial context, memory palace, reinstatement, MVPA

1 Introduction

Many of our memories are intrinsically tied to the locations where they occurred. Thinking about (or actually revisiting) places from our past can immediately bring to mind the meaningful events that occurred there. In this way, our spatial memories can serve as a map not only of physical spaces but also of our remembered experiences in those spaces. In what ways can a spatial context (i.e., the location in which an experience takes place) serve as a scaffold for storing and accessing the details of past episodes? Are there spatial contexts that are more or less effective for attaching event memories, and can we neurally measure the usefulness of a location as a memory cue even before an event has occurred?

Decades of research have found that the representation and retrieval of episodic memories is profoundly tied to spatial location. Prior behavioral research on the context-dependent memory effect suggests that items learned in a particular physical context can be better remembered when the retrieval context matches the encoding context (Smith and Vela, 2001; Godden and Baddeley, 1975), even for contexts that are experienced only through virtual reality (Shin, Masís-Obando et al., 2021) or that are mentally reinstated rather than physically re-experienced (Bramao et al., 2017). Recent behavioral work has also suggested a privileged role for spatial contexts as cues for memory retrieval. For example, spatial context cues: *a)* enhance episodic recall when compared with temporal, thematic (e.g., romantic experience), person, or object cues for imagined or real autobiographical memories (Sheldon et al., 2019; Robin et al., 2016; Sheldon and Chu, 2017; Sheldon et al., 2019); *b)* are spontaneously generated even when not cued by experimenters (Robin et al., 2016; Hebscher et al., 2018), sometimes leading to quicker access to episodic information (Robin et al., 2016; Hebscher et al., 2018, but see Robin et al., 2019); *c)* are associated with richer episodic memory when highly familiar to participants (Arnold et al., 2011; Robin and Moscovitch, 2014, 2017; Robin et al., 2016; Hebscher et al., 2018); and *d)* are associated with preserving long-term recollection of initially low detail memories for both young and older adults (Chang et al., 2024). This behavioral work is complemented by neuroimaging studies of autobiographical memory showing that spatial contexts have a strong influence on the neural representations of remembered or imagined autobiographical events (Robin et al., 2018; Hebscher et al., 2018; Reagh and Ranganath, 2023, among others; for a review see Hassabis and Maguire, 2007). The networks associated with spatial contexts are maintained during multiple phases of memory retrieval, possibly acting as a scaffold for accessing additional event details (Gurguryan and Sheldon, 2019). For example, spatial contexts can be reinstated prior to or concurrently with the retrieval of an item or episode (Herweg et al., 2020; Miller et al., 2013).

Beyond retrieval, prior theoretical work on episodic memory suggests that – at encoding – features of an ongoing experience are bound to the context in which they occur (Yonelinas et al., 2019; Ranganath, 2010; McClelland et al., 1995), allowing spatial contexts to serve as structured "containers" that organize and support the integration of new experiences (Gilboa and Marlatte, 2017; Preston and Eichenbaum, 2013). Consistent with this view, explicitly binding objects to their spatial context during encoding enhances subsequent memory for those objects (Reggente et al., 2020). In another recent study (Shin, Masís-Obando et al., 2021), participants encountered words within two distinct spatial contexts (each associated with a separate schema) and judged each word's relevance to its context without knowing there would be a later memory test. If reinstating the context at retrieval were sufficient to boost memory, all words should have benefitted equally. Instead, only context-relevant words showed a memory advantage, suggesting that these items were more effectively bound to the spatial context during encoding.

Despite the centrality of spatial context in memories, it is unknown whether *(a)* some specific spatial locations are more effective memory cues than others, and if so *(b)* whether this is related to properties of their neural representation. In general, two requirements for robust representation of a memory are thought to be stability over time (allowing for faithful reactivation of the features of the original experience) and distinctiveness (to prevent interference with other similar memories; Sommer and Sander, 2022; Konkle et al., 2010). We hypothesized that these two properties would also be important specifically for building an effective spatial context scaffold – i.e., that spatial locations with more stable and distinctive representations would support better encoding of new information encountered in these locations, and allow for easier access to this information at retrieval. This implies that having a stable and distinctive neural representation for a location *before* associating an object to that location will be predictive of subsequent reinstatement for that object representation.

Our primary mechanistic hypothesis for *why* this would occur was that reliable room representations facilitate the binding of room to object information at encoding (e.g., the sturdier a wall is, the easier it is to hang a painting on it). However, facilitated binding at encoding is not the only way that having a stable and distinctive room representation could facilitate subsequent object reinstatement; an alternative possibility is that having a stable and distinctive room representation has no effect on room-object binding at encoding, and that instead it boosts object recall *indirectly* by boosting the degree to which the room representation is reinstated at test, which – in turn – boosts reinstatement of associated object information (e.g., the brighter the light in a dark room, the easier it is to see what's inside). We will present the results of analyses that control for this alternative possibility.

To test whether reliable spatial contexts scaffold subsequent memory, we custom-built a virtual reality "memory palace" environment of 23 perceptually distinct rooms each with distinct soundtracks, interiors and room-congruent objects, that participants explored using a head-mounted virtual reality display (**Fig 1**). After participants learned the layout of the virtual environment, we used fMRI to compute a neural *room reliability* score for each of the 23 rooms. This score reflected both the stability and distinctiveness of neural representations, measuring the degree to which repeated

presentations of a room evoked patterns that were more similar to each other than to patterns evoked by other rooms. Participants then returned to the VR environment, where they observed (and were asked to memorize) a new salient object that had now been placed into each room. Finally, they performed recall tasks for these items in the fMRI scanner. Overall, our results confirmed our hypothesis: Room reliability, measured before any room-object pairing occurred, predicted the degree of object reinstatement during verbal recall, showing that it is possible to neurally diagnose whether a room will serve as an effective memory scaffold, *before* objects are placed in the room.

2 Results

2.1 Overview

How effective are spatial memory representations as containers for subsequently-bound objects? We sought to answer this question by using the reliability of a pre-learning room representation to predict the degree of reinstatement evidence for recalled objects during self-paced verbal recall. To do this, we needed to quantify **1)** the reliability of a room representation, and **2)** the reinstatement of object information during recall. We defined room reliability as the similarity of a room representation to itself (i.e., stability) minus its average similarity to every other room (i.e., distinctiveness); importantly, this was measured before any room-object associations had been formed (i.e. in the pre-learning phase; see **Fig 2**). Our strategy for quantifying object reinstatement during recall was as follows: We first identified a network of regions involved in the retrieval of objects (the Retrieved Object Classifier Network; ROCN) during a cued-recall task in which participants watched videos of room interiors and were asked to recall the objects that had been randomly assigned to those rooms in VR (see **Fig 4**). We then measured the average classifier evidence for object reinstatement within this network during self-paced verbal recalls, in which participants were instructed to verbally describe with as much detail as possible the rooms and the randomly placed objects in them (see **Fig 5A**). Afterwards, to determine how well the reliability of a pre-learning room representation predicted object reinstatement, we correlated pre-learning room reliability scores with object classifier evidence within the ROCN during self-paced recall trials (see **Fig 5B** and **Fig 6**). We identified a set of regions whose pre-learning room reliability was predictive of object reinstatement during verbal recall, including precuneus, posterior parietal cortex, and prefrontal cortex – specifically, superior frontal gyrus. Importantly, using a model comparison analysis, we also found that some of these regions provided a participant-specific predictive benefit, including posterior parietal cortex, posterior ventral temporal cortex, and superior frontal gyrus (see **Fig 6B**). Lastly, to identify whether room reliability supported object reinstatement *indirectly* by promoting room reinstatement at recall, we conducted a partial correlation analysis controlling for room reinstatement. Even after statistically controlling for room reinstatement, the relationship between room reliability and ROCN object reinstatement remained significant (see **Fig 6C**). Furthermore, no areas showed a significant decrease in the size of this relationship when

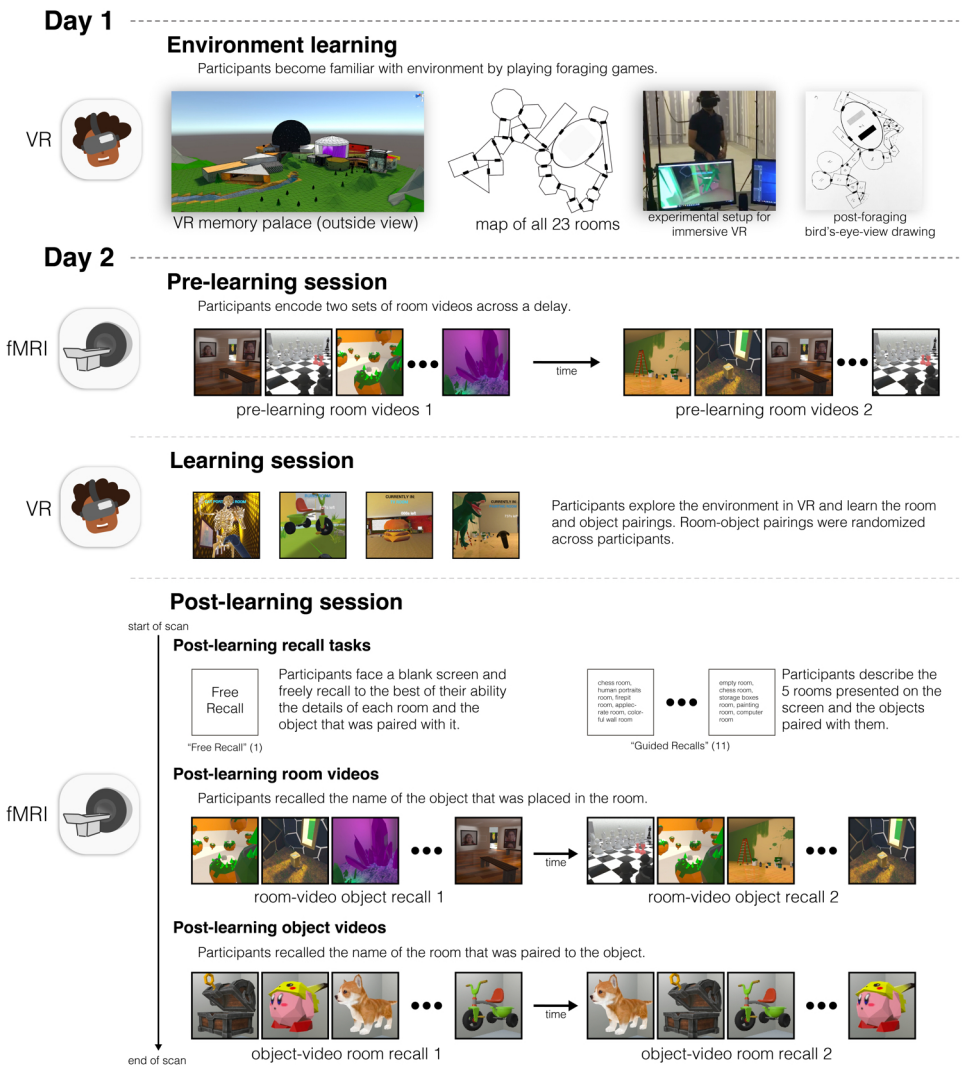

Fig. 1 Experimental paradigm. Participants played a set of foraging games to learn the layout of the 23-room virtual reality environment (photograph of a lab member demonstrating the VR setup used with permission). At the end of each foraging game, to test learning of environment, participants drew a bird's-eye-view map of the environment (see **Fig1 - Supp1** for more examples). 24 hours later, participants were shown room videos in the scanner, with each room presented twice. Participants then re-entered immersive VR, and were given 15 minutes to learn the identities and locations of 23 new objects that had been added to the environment, one per room. Finally, participants returned to the scanner and recalled the items they had seen in a free recall task, a guided recall task (in which they recalled items along specific five-room paths), and a room video task (in which they recalled the item for each presented room). They were also presented with videos of each object, and attempted to recall the room in which each object appeared.

Figure 1–Figure supplement 1. Example bird's-eye-view drawings from a single selected participant.

we controlled for room reinstatement (see **Partial correlation analysis controlling for room reinstatement** Methods section).

2.2 Room reliability

To identify brain regions with reliable room representations for every participant, we compared the similarity of a room's representation across runs to its similarity with representations of other rooms (**Fig 2A**). We ran this analysis on searchlights and hippocampal regions of interest (ROIs; full hippocampus, anterior hippocampus, and posterior hippocampus). We found significant room reliability across most of the cortex. Unsurprisingly, given the audio-visual nature of the room videos, we found high reliability scores in auditory and visual cortex, as well as in precuneus, and posterior hippocampus (**Fig 2C**).

Are there particular room properties, such as size, complexity, or connectedness that contribute to the reliability of room representations? To identify which room features contribute to room reliability, we ran a searchlight analysis where, within each searchlight, we ran a multiple regression predicting room reliability based on six different room features; we generally found that, in default mode network regions, the most reliable rooms tended to be those that were small, had many corners, and had an opening with a view to the outside (**Fig2 - Supp1**).

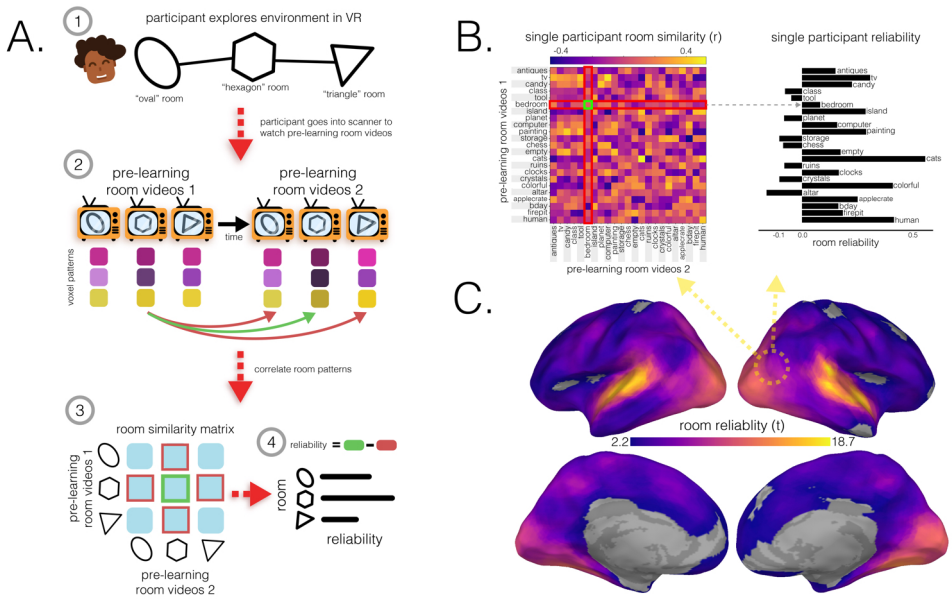

Fig. 2 Room reliability. (A.) Illustration of room reliability methodology. *(1)*: Participants first explored the memory palace in immersive VR and learned its spatial layout by playing a foraging game. *(2)*: In the pre-learning scanning session, before learning room-object pairings, participants watched and listened to videos of each room twice (pre-learning room videos 1 & 2). Each room representation was correlated across runs with every other room, in a searchlight analysis. *(3)*: Room reliability was computed by taking the difference between the similarity of a room pattern to itself (green) and the average similarity of the room with every other room (red). *(4)*: Room reliability was computed for every room, leaving a room reliability score for each of the 23 rooms. The entire procedure outlined in **A** was computed for every participant such that for every searchlight, there were 23 room reliabilities for each of the 25 participants. **(B.)** An example room pattern similarity matrix for one participant, in the searchlight denoted with a dotted circle. This matrix was used to extract room reliability scores as described in **A**, such that for each room (row), the average room similarity to other rooms (red) was subtracted from the room similarity to itself (green). **(C.)** Room reliability across the brain. Colored vertices on the surface indicate regions in which room reliabilities were significantly above zero at the group level ($q < 0.05$), with brighter colors indicating greater reliability.

Figure 2–Figure supplement 1. Relationship between room reliability and room features.

2.3 Behavioral recall

On the second day, participants performed two types of self-paced verbal recall tasks. During the guided recalls (GRs; 11 runs), participants were presented with the names of 5 rooms that followed a path within the virtual palace and were asked to freely recall details of the rooms and the randomly added objects. During the free recalls (FRs), participants were presented with a blank screen and were simply asked to freely recall, in as much detail as possible, the rooms and the added objects. For the guided recalls, we computed accuracy by counting whether a participant recalled the randomly placed objects in that path regardless of whether they were correctly recalled in order

of the path or with the correct room-object pairing. In other words, an object was marked as correctly recalled (out of 5) if it was recalled at any point during the trial. Similarly, for the free recalls, regardless of when an object was recalled, we marked an object as correctly recalled (out of 23) if it was recalled at any point during the free recall. Across both recall types, participants' recalls were at ceiling, with 92% and 80% of participants scoring higher than 90% recall accuracy for guided and free recalls, respectively (**Fig 3D**). We also found that, in both guided and free recalls, participants spent less time speaking about the Empty Room than the across-participant average (**Fig3 - Supp1**) – likely because the room was empty (other than the randomly placed object) and there was less to recall. We also measured the proportion of contiguous room transitions during free recall. Across participants, spatially adjacent rooms were recalled more often than expected by chance ($t(24) = 14.19, p < 0.001$), suggesting an unprompted bias towards contiguous mental traversal (**Fig3 - Supp1F**).

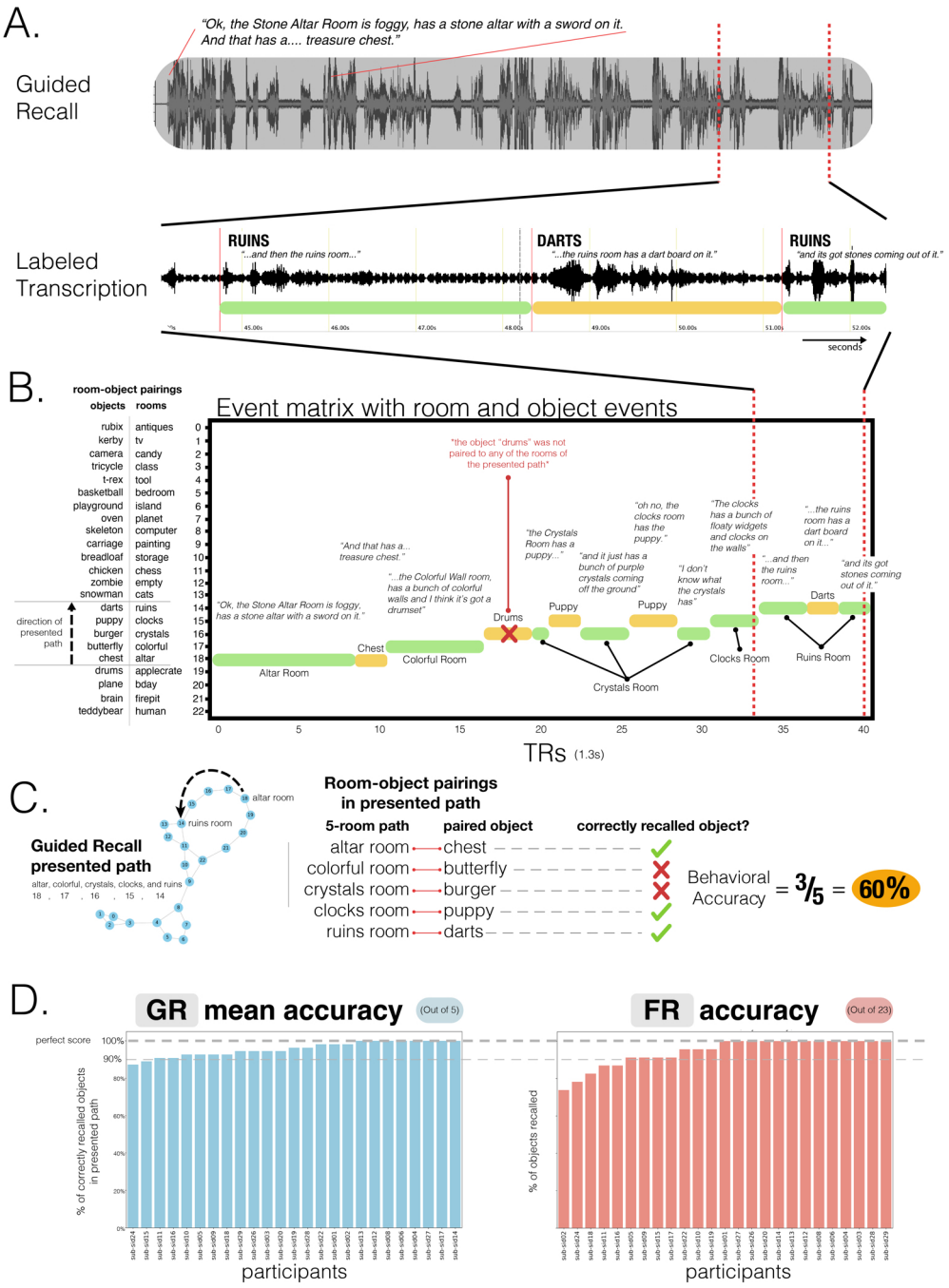

Fig. 3 Behavioral recall scoring. On the second day after learning room-object associations in VR, participants went back into the scanner where they performed a free recall and 11 guided recall tasks. In the free recall task,

Figure 3 continued on next page.

Figure 3 continued.

Fig. 3 participants were asked to recall and describe with as much detail as possible the rooms and the objects paired to them. In contrast, during the guided recall tasks, participants were presented with 5 contiguously connected rooms and asked to describe the rooms and the objects in them. **(A.)** Example guided recall transcription. Participant recalls were transcribed manually for the onset and offset timestamps of when rooms and objects were recalled. **(B.)** Example transcribed recall event matrix. Timestamps of the onsets and offsets of participant recalls were then interpolated from seconds into TRs and organized as event matrices that could then be used to index BOLD recall timeseries. Green and yellow bars indicate room and object recalls, respectively. Object recall timepoints were used to calculate object evidence scores in neural analyses (e.g., the timepoints where participants talked about the Darts were used to measure neural object evidence for recall of the Darts). **(C.)** Guided recall task, pairings, objects recalled and accuracy calculation. First column: This participant was presented with a 5-room path and asked to sequentially describe the rooms and the objects in them. Second column: Calculation of behavioral accuracy for guided recalls. Participants were scored based on whether they recalled the objects that were paired to the rooms in the presented path. Although this participant recalled 4 objects, only 3 were associated with the corresponding cued 5-room path. For both guided and free recalls, points were awarded based on whether participants recalled the relevant objects at any point in time during the recall period, regardless of the order in which the objects were recalled or whether they were recalled in association with the correct room. **(D.)** Guided and free recall accuracy distributions. With these scoring schemes, participants were able to recall objects with high accuracy. Participants' recalls were at ceiling, with 92% and 80% of participants scoring higher than 90% recall accuracy for guided and free recalls, respectively.

Figure 3–Figure supplement 1. Additional behavioral measures.

2.4 Retrieved Object Classifier Network (ROCN)

To measure evidence of object reinstatement during self-paced (guided and free) verbal recall, we first needed to identify a network of regions that represent information about specific objects that were retrieved from memory; to select these regions in a non-circular fashion, we defined these regions using data from room-video object recall trials (**Fig 4A**). In these trials, participants viewed videos of all rooms and verbally recalled the object that had been assigned to each room as it was presented (**Fig 4A**). We used a leave-one-participant-out cross-validation procedure, whereby we made a neural template for each object (using data from a separate phase of the study in which participants viewed object videos) based on object videos from N-1 participants, and then we used these templates to classify the (not-visibly-present) objects being recalled during room viewing in the held-out participant (**Fig 4A**). We opted for this across-participants approach (rather than classifying within-participants) because objects and rooms are confounded within participants, so room information could "leak" into training of a within-participant object classifier; this confound does not exist if training and testing are done across participants, each of whom has their own random set of room-object pairings. In other words, the left-out participant's object templates were never used to classify their own object recall during room videos. We used this procedure to identify the top 50 best object classifier searchlights ($\sim$3% of all searchlights) to make our Retrieved Object Classifier Network (ROCN; **Fig 4B**), which we used as a mask (**Fig 4D**) when measuring object reinstatement evidence during the guided and free recall tasks. We found that the top classifier searchlights were spread throughout cortex and included regions in anterior temporal cortex, frontal gyrus, posterior temporal cortex, posterior medial cortex, and superior parietal cortex, among others (**Fig 4C** and **Fig 4D**). We also conducted additional analyses to extract two

other networks: For one, we classified object patterns while participants watched videos of objects (rather than retrieving object memories) to extract the Perceived Object Classifier Network (POCN), which was entirely, and unsurprisingly, due to the visual task, concentrated in early visual cortex (**Fig4 - Supp1**). For the other, we classified room patterns while participants watched videos of objects (analogous to ROCN, which classified object memories during room videos) to extract the Retrieved Room Classifier Network (RRCN), which was widely distributed and included precuneus, medial prefrontal cortex, anterior temporal cortex and visual cortex (**Fig4 - Supp2**).

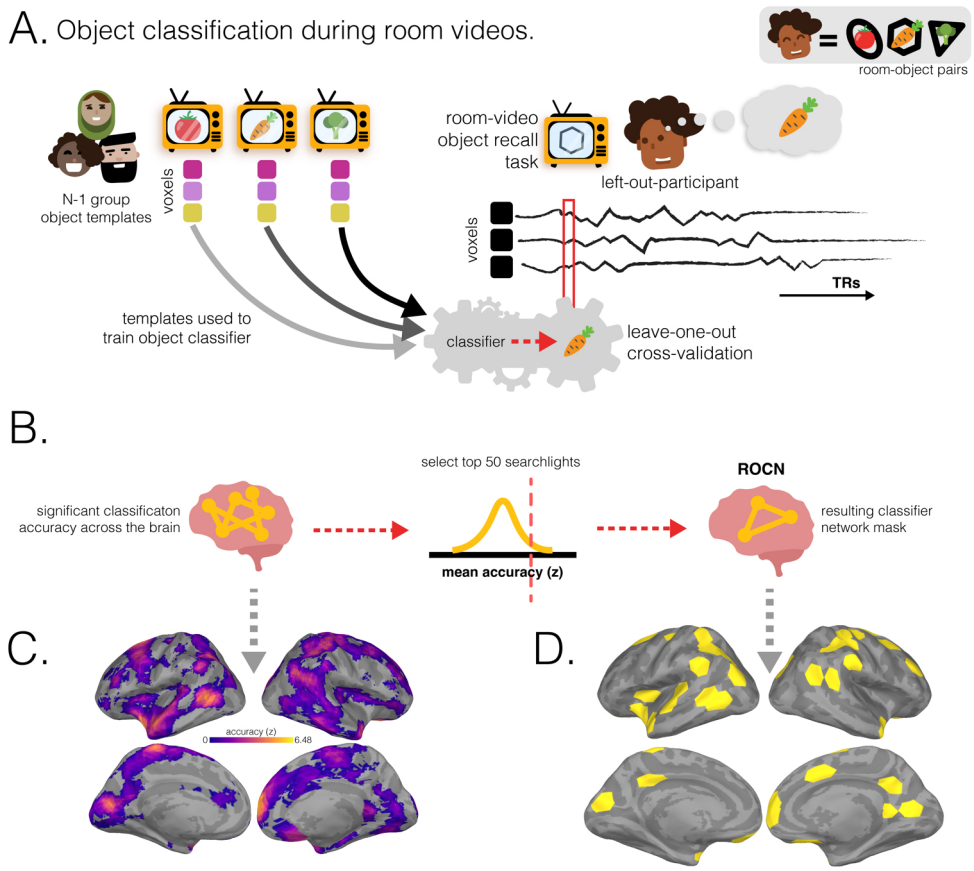

Fig. 4 Retrieved object classifier network (ROCN) methodology and surface maps. **(A.)** During the post-learning room-video object recall task, participants watched a video of a room and verbally recalled the object that was paired to it. In a leave-one-participant-out cross-validation procedure, the characteristic object patterns of the N-1 group – evoked during a separate phase of the study in which participants viewed object videos – were used to train a multinomial logistic classifier. This classifier was then applied to each timepoint on the left-out participant's room-video object recall data. In the pictured example, the left-out participant, Fernando, is recalling the carrot object that was paired with the hexagon room currently being presented. The object classifier, trained on patterns evoked when other participants viewed the objects, was applied to each timepoint of Fernando's recall. We then measured the fraction of timepoints during the hexagon-room video that were classified as activating the carrot representation. **(B.)** For each searchlight, object classification accuracies for both room-video object recall videos for each participant were averaged together and then averaged across participants and z-scored relative to a null distribution. The 50 top-performing searchlights were then selected to form the ROCN. **(C.)** Average object classification accuracy during room-video object recall. The colormap shows the relative classification accuracy across all searchlights (thresholded to show only searchlights with above-chance accuracy). **(D.)** Retrieved Object Classifier Network (ROCN). The top 50 searchlights that were most sensitive to object reinstatement (yellow) were defined as the ROCN for subsequent analyses.

Figure 4–Figure supplement 1. Perceived object classifier network (POCN) methodology and surface maps

Figure 4–Figure supplement 2. Retrieved room classifier network (RRCN) methodology and surface maps.

2.5 Relationship of room reliability and ROCN object reinstatement evidence

Does room reliability predict future object reinstatement during free and guided recalls? Using the object classifier and ROCN searchlights from the previous analysis, we measured the degree of object reinstatement as each participant performed verbal recalls (**Fig 5A**). Note that using neural object reinstatement provided a more sensitive index of successful retrieval than behavioral recall accuracy, since almost all participants were near-ceiling in their retrieval accuracy as described above. Specifically, at each searchlight, we correlated each participant's room reliability with their own composite ROCN object reinstatement score (**Fig6A**; see **Fig6 - Supp3** for an example searchlight). We then averaged these correlations across participants to obtain a searchlight map that we then statistically averaged across recall task types (i.e., guided and free recalls) to get a composite map that indicated regions where room reliability in those regions correlated with subsequent object reinstatement (throughout the ROCN network; **Fig 6A**). Notable positive relationships were observed throughout parietal cortex, prefrontal cortex, superior frontal gyrus, insula, and precuneus. We also found notable negative relationships in right parahippocampal cortex, parts of the motor system, auditory cortex, and ventral visual regions. Importantly, when looking at this relationship separately for guided and free recalls (before generating our composite map), the regions revealed were highly similar, providing an internal replication of this relationship across two categorically different recall task types (**Fig6 - Supp2**)

Lastly, to determine whether room reliability's relationship with object reinstatement was driven by room reinstatement, we ran a partial correlation analysis where we regressed room reinstatement scores in RRCN from both ROCN object reinstatement and pre-learning room reliability, and then correlated the residuals. After controlling for room reinstatement at retrieval, the relationship between room reliability and ROCN object reinstatement evidence remained significant (**Fig 6C**). The pattern of results across the brain shown in **Fig 6C** (when we controlled for room reinstatement) was almost identical to the pattern of results shown in **Fig 6A** (when we did not control for room reinstatement), and there were not any areas where the effect significantly differed between the two maps. Taken together, these results indicate that fluctuations in room reinstatement during retrieval were not responsible for the effects shown in **Fig 6A**. For completeness, we also did this for POCN object reinstatement; similarly to what we found for the ROCN, after controlling for room reinstatement at recall, the relationship between room reliability and POCN object reinstatement remained significant, and there were not any areas where this relationship significantly decreased when we controlled for room reinstatement (see **Fig6 - Supp1C**).

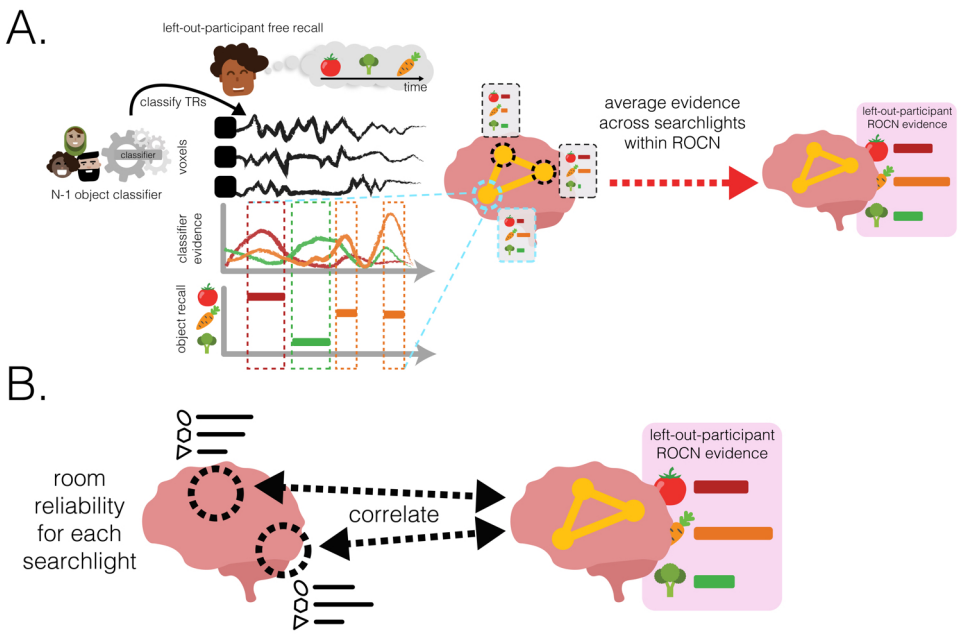

Fig. 5 Methodology for using room reliability to predict object reinstatement (A.) Illustration of methodology for how ROCN object reinstatement evidence was calculated from guided and free recalls. A leave-one-participant-out cross-validation procedure was used with a multinomial logistic classifier to predict object patterns at every timepoint of the left-out participant's recalls. To extract a single composite score of reinstatement evidence within the ROCN for every object and every participant, the classifier evidence for each object recalled was averaged within the ROCN mask across the timepoints when each object was verbally recalled. This yielded a single score for each object and each participant that represented the average object reinstatement evidence in the ROCN during guided or free recall. **(B.)** Illustration of methodology for how object reinstatement evidence was predicted by room reliability. In a searchlight analysis, reliability for a room (in that searchlight) was correlated with the corresponding composite score of reinstatement evidence (within the ROCN mask) for the object paired to that room. This correlation was computed across room-object pairs within each participant and then those correlations were averaged across participants, and finally, across recall task types.

To what extent do the effects in **Fig 6A** reflect group-level differences across rooms (whereby some rooms have both high reliability and high item reinstatement in all participants) versus *participant-specific* differences in which rooms are most reliable in their individual mental maps? To answer this question, we compared the coefficient of determination (R^2) between 1) our original participant-specific model, where each participant's object classifier evidence was predicted using their own room reliability values, and 2) the average R^2 of N-1 models where – in each model – the left-out participant's object classifier evidence was predicted using a different participant's room reliability values (i.e., one model for each of the N-1 other participants). We then took the regions where there was a positive and statistically significant participant-specific effect (i.e., better prediction with the original model), and intersected them with

the correlational analysis performed in **Fig 6A**. This process revealed a participant-specific benefit of room reliability in posterior parietal cortex (near angular gyrus), insula, and superior frontal gyrus **(Fig 6B**). Interestingly, there was also a participant-specific effect where room reliability in a small section of right parahippocampal cortex was negatively associated with ROCN reinstatement evidence.

In a similar fashion to how we related room reliability with object evidence within the Retrieved Object Classifier Network (ROCN), we ran a supplementary analysis in which we quantified object reinstatement within the Perceived Object Classifier Network (POCN; largely composed of visual regions) during verbal recall (**Fig 6 - Supp1**). Across participants, we found generally similar results to the ROCN results, with a positive relationship between POCN reinstatement evidence and room reliability in parietal cortex, superior frontal gyrus, insula, posterior medial cortex, and dorsal occipital cortex. Across both recall tasks, there was a participant-specific benefit of room reliability in posterior parietal cortex, posterior medial cortex, right insula, and portions of right lateral superior and middle frontal gyrus (**Fig6 - Supp1**; refer to **Fig6 - Supp2** for guided and free recalls separately)

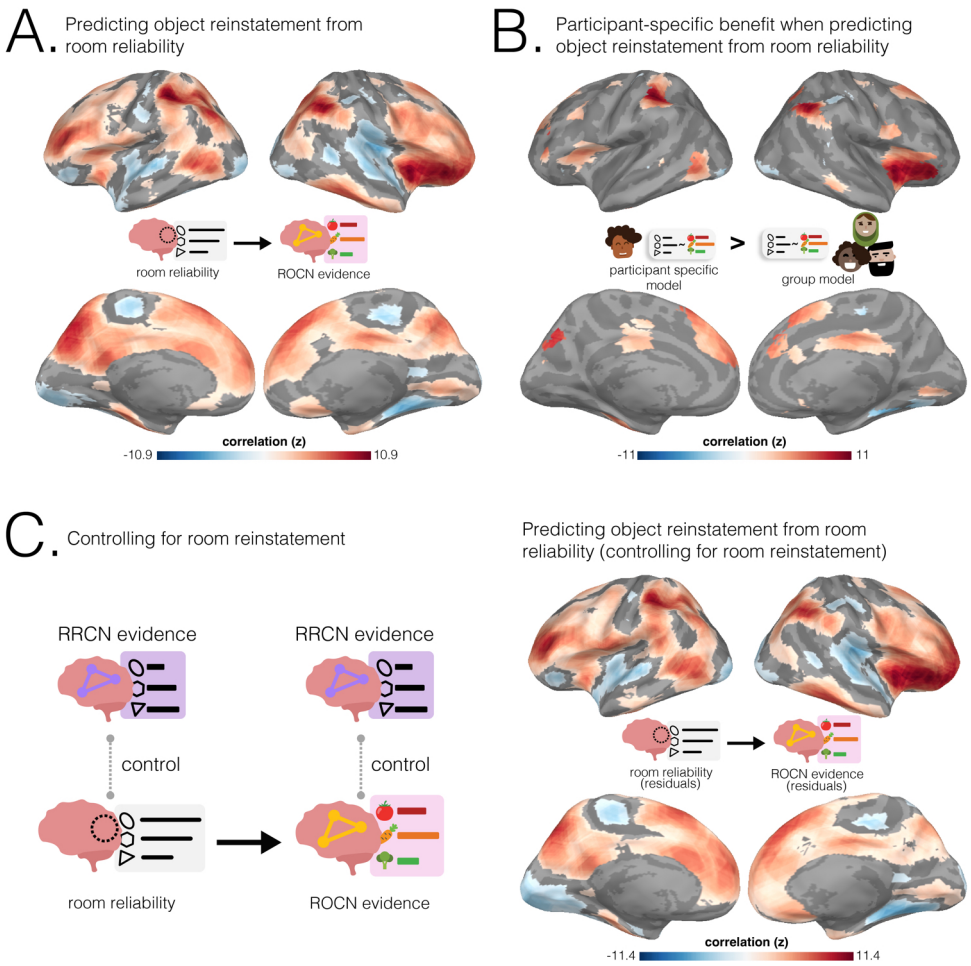

Fig. 6 Predicting ROCN object reinstatement from room reliability. (A.) Relationship between ROCN object reinstatement and room reliability. Regions where room reliability predicted ROCN object reinstatement across both guided and free recalls. Objects placed in rooms with the most pre-learning neural stability in these regions were reinstated the most strongly during retrieval. **(B.)** Model comparison results. Regions where room reliability predicted ROCN object reinstatement across both guided and free recalls and there was a predictive benefit from participant-specific room reliability. In these regions, the rooms that were most reliable for a specific participant (rather than rooms that were generally reliable across the group) were predictive of object recall for that specific participant. The surface maps presented in **B** show the intersection of the participant-specific models shown in **A** and the regions where there was a significant positive difference in the coefficient of determination between the original participant-specific model and the N-1 group model. Statistical significance for the differences between the coefficients of determination was determined by comparing the differences to a null distribution and FDR-correcting for $q < 0.05$.

Figure 6 continued on next page.

Figure 6 continued.

Fig. 6 (C.) Controlling for room reinstatement. Left column: Schematic illustrating how room reinstatement evidence in RRCN (during timepoints in which participants verbally recalled a room or its paired-object) was regressed out of room reliability and ROCN object reinstatement scores. Room reliability residuals were then correlated at each searchlight with ROCN object reinstatement residuals. Right column: Regions where room reliability predicted ROCN object reinstatement after controlling for room reinstatement during room and object recall. The surface maps presented in **A** and **C** were statistically thresholded by comparing correlations to a null distribution and then FDR-correcting for $q < 0.05$. All three surface maps are colored based on the magnitude of the z-scored correlation values of the participant-specific model, with blue showing negative and red showing positive relationships, respectively.

Figure 6–Figure supplement 1. Predicting POCN object reinstatement from room reliability

Figure 6–Figure supplement 2. Predicting ROCN and POCN object reinstatement from room reliability for guided and free recalls.

Figure 6–Figure supplement 3. Correlation plots between room reliability and object reinstatement in ROCN for an example searchlight.

3 Discussion

In this study, we posited that a cognitive map of spatial contexts is most useful as a container for future memories when locations have reliable representations, providing specific and consistent cues every time they are accessed. To test how the neural properties of a spatial context memory supports new memories, we developed a novel paradigm that allowed us to quantify the within-participant reliability of a spatial context memory, before it became the location in which a new memory was formed, and then used it to predict the degree at which that new memory was remembered. We did this by having participants develop spatial context memories of a 23-room immersive virtual reality memory palace, scanning them to extract the neural properties of their spatial memories for "empty" rooms within the palace (pre-learning phase), and then scanning them again afterwards, as they verbally recalled the "filled" rooms and the objects that filled them (post-learning phase). For the first time, we were able to show that pre-learning room reliability – the representational quality of an "empty" memory scaffold – is predictive of post-learning object reinstatement in two types of verbal recall; we further demonstrated that – in some regions – a participant's idiosyncratic room reliability values provided a predictive advantage, beyond what could be predicted by knowing (in general) which rooms were most reliably represented across participants. Finally, we showed that this relationship between room reliability and object reinstatement persists even after statistically controlling for room reinstatement at recall. By ruling out the alternative hypothesis that fluctuations in room reinstatement are (fully) driving the effect, this control analysis provides indirect evidence in support of our preferred hypothesis – namely, that reliable room representations scaffold memory for objects by facilitating the binding of objects to rooms at encoding.

Theories in cognitive psychology have long argued that we develop knowledge structures that help to organize new information during encoding and later serve as a scaffold to recall specific details (Graesser and Nakamura, 1982); for example, prior work has discussed how event schemas (Schank and Abelson, 1977), which describe the

protoypical sequence of events associated with well-learned experiences (e.g., restaurant visits), can support memory for new life events. In a similar fashion, knowledge about the structure and affordances of a spatial context can scaffold memories for experiences that occur in that context (Robin et al., 2016; Brunec et al., 2018). Our results support this general framework but also argue that all schematic containers are not equally effective at organizing memories; contexts that are only weakly learned and/or suffer interference from other contexts will not be effective scaffolds, consistent with work showing that repeated exposure to a single room versus distributed exposure to many rooms creates a more effective contextual cue (Smith, 1979). Additionally, our findings here also provide further support on the utility of VR as a tool for studying how spatial contexts can shape memory and behavior (Reggente et al., 2018).

3.1 Room reliability is predictive of object reinstatement

There are two important features that make this study uniquely placed to investigate the role of spatial context scaffolds in episodic memory. First, the virtual rooms in this study are experienced in immersive VR and vary widely along many dimensions (room size and geometry, decoration, background soundtrack, etc.), allowing participants to create rich and unique representations of individual rooms. Second, unlike other studies, neural patterns for each of the spatial contexts were acquired before the key learning event took place (here, the newly placed object in each location). These two features provided us with the opportunity to relate the neural patterns for "empty" spatial contexts with the reinstatement of the objects that had been placed in them in a subsequent part of the experiment.

Specifically, our paradigm allowed us to relate the reliability of a room representation (the "empty" scaffold) across cortex to the reinstatement of the objects that had been placed in rooms explored in VR. In general, we found that object reinstatement was predicted by room reliability in precuneus, insula, frontal cortex, and regions throughout lateral parietal cortex (**Fig 6**) suggesting that measuring the structural integrity of a spatial context representation before a life-episode is predictive of how well that episode will be reinstated later. Moreover, these effects were found separately for both guided and free recall, providing an internal replication of our results and suggesting that stable context representations are useful for retrieval across multiple kinds of memory tasks. We observed strong effects in regions that are well-known to support mental and virtual navigation (Ghaem et al., 1997; Maguire et al., 1997; Hartley et al., 2003; Brodt et al., 2016; Rosenbaum et al., 2004; Ino et al., 2002; Epstein and Baker, 2019; Spreng et al., 2009; Epstein et al., 2017), including precuneus and the dorsal occipital lobe. Similar regions have also been identified in many types of tasks involving spatial knowledge: during spatially-cued retrieval of real or imagined autobiographical memories (Robin et al., 2018; Gurguryan and Sheldon, 2019; Szpunar et al., 2014), during recognition or retrieval of the spatial context in which an item was encountered (Burgess et al., 2001; Cooper and Ritchey, 2019; Hayes et al., 2004), during the recollection of spatial relationships in 2D and 3D (Frings et al., 2006; Hirshhorn et al., 2012; Schott et al., 2018; Dordevic et al., 2022), during reinstatement of spatial contexts during item retrieval (Essoe et al., 2022), and during the encoding

and retrieval of items bound to a spatial context (Kondo et al., 2005; Flanagin et al., 2023).

Although these studies highlight the importance of spatial knowledge in a diverse range of learning and memory tasks, most of these studies focused on univariate or functional connectivity changes during the tasks, with few leveraging multi-voxel pattern analyses (e.g., Robin et al., 2018; Essoe et al., 2022; Huang et al., 2024), and none quantifying the quality of the specific spatial representations used in these tasks. Thus, our work here in combination with these prior studies, adds to the vast literature on spatial memory, and provides a potential prerequisite for the successful completion of any spatial task: Spatial context representations need to be reliable to be useful for subsequent memory storage.

In some other brain regions, we observed that room reliability in those regions was negatively related to subsequent object reinstatement. How can we explain these negative relationships? Since these regions are primarily in lower-level auditory and visual cortex, one possibility is that these regions code for lower-level sensory features, not spatial contexts, and the room reliability observed in these regions was actually a measure of how strongly these sensory properties were being represented. In this case, stronger representation of isolated features could be at odds with larger-scale and gist-like representations of the room geometry and semantic properties, making a room less useful as a contextual anchor for subsequent object memory. Further work investigating object representation in the brain and its relationship to room reliability is required to aid in parsing the negative relationships we found.

What underlying mechanisms explain the relationship between object reinstatement and room reliability? Our hypothesis was that reliable room representations scaffold memory for objects by facilitating the binding of objects to rooms at encoding. Our finding that room reliability (measured prior to encoding) correlates with object reinstatement (measured during recall) is compatible with this "facilitated binding at encoding" hypothesis. An alternative possibility is that successful reinstatement of reliable rooms during recall promotes object reinstatement for these rooms; this could give rise to a correlation between room reliability and object reinstatement, even in the absence of facilitated binding at encoding. We addressed this alternative hypothesis by controlling for room reinstatement during verbal recall, and found that the relationship between room reliability and object reinstatement remained significant; furthermore, there were not any areas that showed a significant decrease in the size of the effect when we controlled for room reinstatement. The results of this control analysis provide indirect support for our hypothesis that room reliability supports improved room-object binding at encoding; namely, a reliable spatial context representation may provide a stable schematic map that facilitates the integration of new episodic content – the more reliable the container, the easier it is to populate it with information. Future work in which participants are scanned during object-location encoding would help shed additional light on how room reliability enhances the creation of episodic memories.

3.2 Room reliability

We described the representational stability and distinctiveness of a spatial context through a reliability score that measured the specificity of a room's representation across runs. These spatial contexts were designed to be visually and auditorily rich to reflect real-world contexts. Given that room reliability was derived from audio-visual stimuli, it was not surprising to find the strongest reliability in visual and auditory cortex. In addition to these sensory regions, we found significant room reliability in other regions that have been implicated in higher-level processing: parietal cortex (including intraparietal sulcus), posterior medial cortex (including precuneus), and lateral prefrontal cortex (including premotor cortex). In other studies, these regions have been shown to maintain specific scenes or events within stories along various time scales during movie watching (Baldassano et al., 2017; Hasson et al., 2015, 2008; Lerner et al., 2011; Masís-Obando et al., 2022). These regions may help to ensure stable and distinctive representations of the high-level properties of the current situation that go beyond low-level sensory properties — an idea consistent with prior work showing that these regions represent event-types shared across stories, regardless of whether the story is presented as an audio narrative or an audiovisual movie (Baldassano et al., 2018; Masís-Obando et al., 2022). Although some of this event structure can arise from the temporal dynamics of the stimulus itself, internal schemas can also be used to actively organize an experience into stable events (De Soares et al., 2023). Our results suggest that this kind of top-down stabilization may be most effective when the schema itself is highly reliable, providing a robust starting point for building episodic event representations.

Although high pattern similarity across identical trials is related to better subsequent memory (Xue et al., 2010), purposefully increasing variability in item encoding by varying the encoding context has been shown to improve item memory (Salan et al., 2024; Sievers et al., 2019), perhaps by increasing the number of possible retrieval cues for the item (see, e.g., Melton, 1970; Lohnas et al., 2011). It is therefore possible that there are some situations in which *un*stable context representations would be useful for creating memories, e.g., if items are studied multiple times in a context and then recognition memory is tested in a novel context. However, in our paradigm, participants were explicitly using a context-based strategy for retrieving items, mentally simulating rooms and trajectories through rooms in order to reinstate item memories. In this case, we would expect that having a reliable contextual index for episodic memories would be critical for effective recall of items, consistent with our findings that stability in scene-related brain regions predicted item reinstatement. Future work could investigate whether this relationship disappears or reverses in other situations, such as when many items are paired with the same room (reducing the usefulness of rooms as memory cues), or when rooms have features that vary, e.g., with time of day (such that representational variability might reflect meaningful changes in contextual features), or when the recall task requires reporting only objects while suppressing recall of room features. Similarly, novelty may influence how room reliability scaffolds memory: a new context may be less stable than a highly-familiar location, but could

still enhance memory because its novelty promotes additional attention and processing. Future work examining how repeated exposure and contextual novelty interact with reliability could shed new light on their contributions to memory.

3.3 Our experimental paradigm and the method of loci

Our "memory palace" paradigm draws inspiration from the mnemonic technique called the method of loci (MOL), in which items are associated with an *imagined* sequence of spatial locations in a pre-learned map. However, our study diverges from this technique in several key ways: Unlike many implementations of MOL, participants were not required to encode or recall to-be-remembered items in an explicit linear sequence of rooms, nor were they instructed to use any particular mnemonic during room-object binding. Instead, participants explored the virtual environment freely and developed their own strategies for memorization.

Despite these differences, the motivation for this technique is related to the hypothesis tested in this study: that a well-learned spatial map consisting of many distinct locations is the optimal encoding environment for new item memories. The learnability of this technique suggests that it may rely on inherent spatial memory structures shared across people. In fact, the ability to improve memory through this spatially-based technique has been shown across multiple studies behaviorally and neurally (behavioral: Reggente et al., 2020; Legge et al., 2012; Roediger, 1980; Bower and Reitman, 1972; Moè and De Beni, 2005, among many; neural: Dresler et al., 2017; Wagner et al., 2021; Nyberg et al., 2003; Maguire et al., 2003; Mallow et al., 2015; Kondo et al., 2005; Fellner et al., 2016; Liu et al., 2022; Huang et al., 2024). Generally, neuroimaging studies of this technique have largely focused on the impact of MOL (at varying levels of training or compared to other mnemonics) during item encoding (Maguire et al., 2003; Nyberg et al., 2003; Dresler et al., 2017; Wagner et al., 2021), with only a few performing univariate contrasts during recall (Kondo et al., 2005; Mallow et al., 2015; Fellner et al., 2016; Liu et al., 2022), and only one, to our knowledge, examining multivariate pattern activity for loci, items, and their conjunctive associations (Huang et al., 2024). The univariate results during recall have shown enhanced engagement of regions including retrosplenial cortex and precuneus after instruction in MOL (Kondo et al., 2005), suggesting that spatial representations of loci are strategically activated during retrieval. A recent study measuring multivariate activity patterns during MOL (Huang et al., 2024) found robust representations for individual loci during the creation and retrieval of item-locus pairs in regions including precuneus and posterior parietal cortex, suggesting potentially overlapping mechanisms in how our naive subjects and MOL-trained individuals use spatial information for item memorization. It remains an open question whether enhanced room reliability helps support memorization when using MOL.

4 Conclusion

After participants learned a complex spatial environment in VR, we measured the neural reliability of each spatial location within this map. When participants later used this environment to learn a new set of room-object associations, we showed that

this room reliability measure could predict the degree to which objects associated with each room successfully came online during naturalistic recall. Together, these results showcase how the quality of a spatial context can be quantified and used to 'audit' its utility as a memory scaffold for future memory.

5 Methods & Materials

5.1 Participants

Data were collected from a total of 30 participants between the ages of 21–32 (16 females, 14 males) with normal or corrected-to-normal visual acuity. At the end of the study, participants were paid and debriefed about the purpose of the study. Every effort was made to recruit an equal number of female and male participants and to ensure that minorities were represented in proportion to the composition of the local community. The experimental protocol was approved by the Institutional Review Board (IRB) of Princeton University, and all participants provided their written informed consent (IRB #7225). Due to technical difficulties (corrupted and missing files), 5 participants were excluded, leaving a total of 25 participants (11 females, 14 males).

5.2 Stimuli

5.2.1 Virtual reality stimuli

Environment: A custom-built virtual reality environment made up of 23 interconnected distinct rooms with distinct soundtracks was explored by participants using a head-mounted virtual reality display. Each of the rooms was built to be as visually and aurally distinct as possible. To that end, for visual distinctiveness, each room followed a different theme (e.g., planetarium, computer store, etc) with theme-congruent objects carefully placed throughout, and the rooms had different shapes (e.g., oval, rectangle, etc) and different sizes (e.g., large, small, etc). To promote auditory distinctiveness, each room had a distinct soundtrack on loop that was audible only when a participant entered each room and some rooms contained specific sound effects that matched the room context (e.g., bird chirps if the room had a window facing the outside).

The majority of rooms were connected with only 2 other connecting rooms, while a few, 'hubs', had more than 2 connecting rooms. Of all 23 rooms, 16 rooms (70%) were connected with 2 other rooms, 6 rooms (26%) were connected with 3 other rooms, and 1 room (4%) was connected with 4 other rooms.

To reduce the potential for motion sickness, participants explored the virtual world while seated in a 360 degree rotatable chair, and any instance of participant-initiated teleportation was followed by a short and smooth fade-in-and-out of black. Participants teleported within and between rooms by pressing a button on a wireless controller that would appear digitally reconstructed in VR as a 3D object. The range of teleportation was limited to force teleportation across small distances and to avoid fast teleportation across rooms. Rooms were connected by doorways; given the current room a participant was in, only the immediately connected rooms were visually accessible via the doorways, while further-away rooms were culled from view.

Music and Sounds: Sounds of birds, ambience, firewood crackling, and others, were manually recorded or freely downloaded from the internet. Music for each room was either custom-composed in Ableton Live software, downloaded from the internet, or requested from professional composers (**Supplementary Table 1**).

All tasks were presented on a wired HTC Vive head-mounted display (1,080 × 1,200 resolution per eye, with a 90-Hz refresh rate, and built-in headphones and integrated microphone), which was connected with a wire to a computer running 64-bit Windows 10 on an Intel Core i7-6800K CPU @ 3.40GHz with 32GB RAM and an Nvidia GeForce GTX 1080 graphics card.

All tasks and visual presentations were created and coded in Unity3D 5.5.2f1 (and 2017.1.2f1), a game-development platform, with Virtual Reality Toolkit (VRTK; vrtk.io), a virtual-reality programming tool-kit for Unity3D. The majority of 3D models, textures, environments and other assets were custom-built using SketchUp (sketchup.com) or Blender (blender.org). The remaining assets were downloaded from the Unity Asset Store (assetstore.unity.com), Turbosquid (turbosquid.com), or other publicly available online repositories, and then modified using Blender to reduce model complexity and size.

5.2.2 Scanning stimuli

During scanning, participants were presented with videos of rooms and videos of objects. These videos were generated beforehand and presented to participants in a pseudorandom order.

Room videos: To generate the room videos using Unity, a virtual camera was placed in the center of each room. The camera was scripted to rotate a full 360 degrees to capture the panorama of each room within 10s. OBS Studio (obsproject.com) was used to screen capture the output of the virtual camera. Each room video lasted 10s and was followed by a 5s interstimulus interval before the next video.

Object videos: To generate the object videos, a virtual photography studio was created with a blank backdrop and a 3 point lighting setup. All 23 objects were placed in the center of the virtual studio and scripted to rotate 360 degrees in front of a virtual camera facing them within 10s. OBS studio was used to screen capture the output of the virtual camera. Similarly to the room videos, each object video lasted 10s and was followed by a 5s interstimulus interval before the next video.

Stimulus presentation: All generated stimuli were presented to participants in the scanner using PsychoPy (Peirce, 2007) to time task and stimulus presentations with the scanner trigger. Every presented video or task instruction was preceded by a 5s black screen.

5.3 Data acquisition and pre-processing

5.3.1 MRI acquisition & pre-processing

MRI data were collected on a 3T full-body scanner (Siemens Prisma) with a 64-channel head coil. Functional images were acquired using an interleaved multiband EPI sequence (TR = 1300ms, TE 33ms, flip angle 80 degrees, whole-brain coverage, 2 mm

slice thickness, FOV 192 mm^2, SMS = 4). Anatomical images were acquired using a T1-weighted magnetization-prepared rapid-acquisition gradient echo (MPRAGE) pulse sequence (1 mm^3 resolution). Anatomical images were acquired in a 6-min scan before the functional scans; during this scan, participants watched videos of paragliding from YouTube. Field maps were collected but not used in our preprocessing pipeline.

All raw data acquired from MRI were converted to BIDS formatting (BIDS version 1.0.1), anatomical images were de-faced using pydeface (version 2.0.0), and resulting data were subsequently preprocessed using fMRIPrep version 1.0.3, a Nipype (Gorgolewski et al., 2011, 2017) based tool. Each T1w (T1-weighted) volume was corrected for INU (intensity non-uniformity) using N4BiasFieldCorrection v2.1.0 Tustison et al. (2010) and skull-stripped using antsBrainExtraction.sh v2.1.0 (using the OASIS template). Brain surfaces were reconstructed using recon-all from FreeSurfer v6.0.0 (Dale et al., 1999), and the brain mask estimated previously was refined with a custom variation of the method to reconcile ANTs-derived and FreeSurfer-derived segmentations of the cortical gray-matter of Mindboggle (Klein et al., 2017). Volume-based spatial normalization to the ICBM 152 Nonlinear Asymmetrical template version 2009c (Fonov et al., 2009) was performed through nonlinear registration with the antsRegistration tool of ANTs v2.1.0 (Avants et al., 2008), using brain-extracted versions of both T1w volume and template. Brain tissue segmentation of cerebrospinal fluid (CSF), white-matter (WM) and gray-matter (GM) was performed on the brain-extracted T1w using fast (Zhang et al., 2001) (FSL v5.0.9). Surface-based normalization based on nonlinear registration of sulcal curvature was applied using the fsaverage6 surface template from FreeSurfer.

Functional data was slice time corrected using 3dTshift from AFNI v16.2.07 (Cox, 1996) and motion corrected using mcflirt (FSL v5.0.9 Jenkinson et al., 2002). "Fieldmap-less" distortion correction was performed by co-registering the functional image to the same-participant T1w image with intensity inverted (Huntenburg, 2014; Wang et al., 2017), constrained with an average fieldmap template (Treiber et al., 2016), implemented with antsRegistration (ANTs). This was followed by co-registration to the corresponding T1w using boundary-based registration (Greve and Fischl, 2009) with 9 degrees of freedom, using bbregister (FreeSurfer v6.0.0). Motion correcting transformations, field distortion correcting warp, BOLD-to-T1w transformation and T1w-to-template (MNI) warp were concatenated and applied in a single step using antsApplyTransforms (ANTs v2.1.0) using Lanczos interpolation.

Physiological noise regressors were extracted applying CompCor (Behzadi et al., 2007). Principal components were estimated for the two CompCor variants: temporal (tCompCor) and anatomical (aCompCor). A mask to exclude signal with cortical origin was obtained by eroding the brain mask, ensuring it only contained subcortical structures. Six tCompCor components were then calculated including only the top 5% variable voxels within that subcortical mask. For aCompCor, six components were calculated within the intersection of the subcortical mask and the union of CSF and WM masks calculated in T1w space, after their projection to the native space of each functional run. Frame-wise displacement (Power et al., 2014) was calculated for each functional run using the implementation of Nipype.

5.3.2 Additional pre-processing

After fMRI data were aligned and preprocessed to fsaverage6 resampling, the resampled data were further preprocessed with a custom Python script that removed nuisance regressors which included the 6 degrees of freedom motion correction estimates; framewise displacement: the estimated bulk-head motion; head motion estimates from white matter and CSF, and the cosine bases for high pass filtering to account for low-frequency signal drifts (up to 0.008 Hz, or 125 seconds). Within the same python script, the resulting timeseries data was z-scored for each run (i.e,. task), such that there was a single preprocessed timeseries per task (e.g., pre-learning room videos, post-learning object videos, recall, etc).

5.4 Experimental paradigm

The study took place on two consecutive days and was composed of a behavioral session on day 1 and a behavioral and two scanning sessions on day 2.

Day 1

On day 1, participants were familiarized with the virtual environment and exposed to two VR foraging games and hand-drawing tasks to facilitate the learning of the spatial layout. Specifically, on day 1, after participants read and signed the consent and screening documents, participants were informed about what they would be experiencing in VR and about the safety measures taken to ensure their safety and comfort. They were told that they would be seated to decrease potential dizziness that arises more commonly during VR that involves standing. They were also informed that at any time the experiment could be stopped if they are feeling uncomfortable or dizzy. They were told that they would play two foraging games in virtual reality that involve freely moving through the virtual reality environment with the goal of collecting floating cubes. In the first game they had to collect a cube from every room. In the second game they had to repeatedly navigate to designated rooms to collect additional cubes. They played the second game twice. Between each game, participants were asked to draw a bird's-eye-view map based on their current knowledge of the environment (**Fig1 - Supp1**) – we did this to ensure participants were learning the spatial layout of the environment. By the end of the behavioral session participants had completed a total of three games and three maps. Throughout the experimental session, the experimenter asked the participant about their overall comfort and reminded them if they feel dizzy or nauseous, the experiment could be easily stopped without consequence. After the completion of the foraging tasks, the participants were compensated and reminded to return the next day for the two scanning sessions and the additional VR behavioral session.

Day 2

On Day 2 (1 day later), three sessions took place: In the first session, participants were scanned with fMRI for a small battery of encoding tasks (pre-learning scan); in the second session, participants learned room-object associations in VR for randomly placed objects in each of the 23 rooms (learning behavioral session), and in the third session, participants were scanned again with fMRI as they proceeded through a battery of encoding and retrieval tasks (post-learning scan).

Session 1 (pre-learning scan): On day 2, participants were greeted at the MRI room, asked to draw a bird's-eye-view map of the environment (as had been done the day before). After listening to a short unrelated audio clip in the scanner to verify volume level, participants were told that they would be presented with two sets of audiovisual stimuli of the rooms. In the first set they saw 360-degree room rotation videos of all the rooms (i.e., pre-learning room videos) and were instructed to verbally recall the name of the room when they recognized it. The second set, which was viewed after the first, was exactly the same as the first, except room order was randomized for each participant. Every stimulus presentation was preceded by a 5s blank screen.

Session 2 (learning behavioral session): After participants finished the pre-learning scan, they were taken out of the scanner bore and instructed to carefully stand up. They were then guided back to the behavioral room with the VR equipment to complete the second session of VR. In this session, participants were refamiliarized with the environment by playing the first foraging game again. Afterwards, they drew a bird's eye view map once again and then were told that when they returned to the virtual world, they would find 23 different 3D objects scattered in each of the 23 rooms. They were then given 15 minutes to memorize the room-object pairings.

Session 3 (post-learning scan): After the 15 minutes that participants were given to memorize the room-object pairings had elapsed, participants were guided back into the MRI room. Before getting into the scanner, participants were told that they would be asked to verbally recall in as much detail as possible the 23 room-object pairings. They were also told that they would be presented with the same audiovisual stimuli from Session 1, and they would also view an additional set of videos that included objects. In the first task (Free Recall), participants were asked to describe in as much detail as possible all the rooms and objects that they saw in VR. In the second task (Guided Recall), participants were asked to recall with as much detail as possible the appearance of the rooms and objects along specific 5-room paths within the environment. The names of the 5 rooms were visible on screen. They did this guided recall task 11 times, each time with a different 5-room path. When they had completed recalling the rooms and objects to the best of their ability for the Free Recall and Guided Recall tasks, they were told to inform the experimenter by saying "Done". In the third task (which we label as room-video object recall), participants were exposed to the same 360-degree room rotation videos from the aforementioned pre-learning room video tasks, but this time when they were shown a room video, they were tasked to recall the novel object that had been placed in it (i.e., room-video object recall). They did this task twice for all rooms. Because room-object pairings were generated randomly for each participant, the objects recalled during this task were usually different across participants. Afterwards, in the fourth in-scanner task, participants saw the post-learning object videos. During these, participants performed the object-video room recall tasks: Participants were shown 360-degree object rotation videos and instructed to say the name of the room that was paired with that object. They did this task twice for all objects.

5.5 Searchlights

Our searchlights were generated by constructing them with every valid vertex as their center, then iteratively removing the most-redundant searchlights until no more could be removed while covering each vertex with at least 10 searchlights. This process yielded 1483 searchlights per hemisphere.

5.6 Hippocampus

Our full hippocampus region of interest (ROI) was extracted from a freesurfer subcortical parcellation. This ROI was then split into an anterior portion ($y > -20$) and posterior portion ($y \leq -20$) in MNI space (Guo and Yang, 2020; Poppenk et al., 2013; Masís-Obando et al., 2022).

5.7 Behavior

5.7.1 Behavioral event matrices

Pre-learning and post-learning room, and object videos: The timing of stimulus presentations for every room and object was logged, and a custom python script was used to convert the timestamps to a behavioral timeseries event matrix that marked the start and end of every stimulus presentation for every participant. The resulting matrix that contained the timing (in milliseconds) and room or object identity was then downsampled to 1.3s TRs and used in subsequent analyses to index into a participant's BOLD timeseries data to identify the moments in time participants were encoding a specific video. In sum, the python script generated 6 different behavioral event matrices, 2 pre-learning room event matrices (i.e., pre-learning room videos), 2 post-learning room event matrices (i.e., room-video object recall tasks), and 2 post-learning object event matrices (i.e., object-video room recall tasks).

Post-learning free recall and guided recall: Participants were asked to recall and describe the rooms in the virtual environment and the objects paired to the rooms. Using TotalRecall (memory.psych.upenn.edu/TotalRecall) audio files of participant's recalls were imported and transcribed by timestamping the start of a room or object verbal description. For example, if a participant said, "I remember walking through the chess room, it had large chess pieces. The object in there was a basketball..." the start and end of the "chess room" timestamps would have been at the start and end of the first sentence, respectively. This is because we assumed that the room would have come to mind at the start of the sentence rather than midway. Similarly, the object start timestamp would have been considered the start of the second sentence. For every participant, these timestamps were then imported into a custom Python script that generated a behavioral timeseries event matrix that marked the start and end of each verbal room or object recall. This resulted in 11 guided recall and 1 free recall behavioral timeseries event matrices that indicated the trajectory of room or object recalls. These were then downsampled to 1.3s TRs and used in subsequent analyses to index into a participant's BOLD timeseries data to identify the moments in time a participant was recalling a particular room or object.

Time spent recalling rooms or objects: To assess whether certain rooms or objects were discussed significantly more than others during recall, we conducted an across-participant global mean comparison. For each room and object, we computed the mean time spent speaking across participants. We then performed a one-sample t-test for each room (or object), testing whether its average recall time significantly deviated from the grand mean (i.e., the average across all rooms or objects). Next, we applied a Bonferroni correction to the resulting p-values to account for multiple comparisons.

Contiguity in free recall: To assess whether participants tended to recall spatially connected rooms in sequence, we computed, for each participant, the proportion of times each room transition was to an adjacent room (i.e., graph distance of 1 in the adjacency matrix of the virtual environment). Self-loops, where the same room was recalled consecutively, were excluded. To calculate the baseline probability that a participant may have recalled an adjacent room just by chance, for each transition, we counted the number of currently adjacent rooms divided by the 22 possible other rooms (excluding the current room) and then averaged across all transitions. To test for significance, we ran a paired sample t-test where we compared each participant's proportion of contiguous recall with their chance baseline (see **Fig3 - Supp1F**).

5.8 fMRI analysis

5.8.1 Characteristic object patterns ("object templates")

To acquire the characteristic neural patterns for objects ("object templates") we created 23 regressors to model the neural response to each of the 23 objects. We placed each of the 23 object regressors in a design matrix that marked the transitions between object videos across both post-learning object video tasks; the matrix was convolved with a hemodynamic response function (HRF) from AFNI (Cox, 1996) and then z-scored. We then extracted the characteristic spatial pattern across vertices for each object by fitting a general linear model (GLM; within each participant) to the time-series of each vertex using these 23 regressors. Doing this simultaneously across both post-learning object videos yielded a single set of 23 characteristic object spatial patterns across vertices for each participant. These object templates, which were obtained for every participant, were then used in subsequent analyses for training multinomial logistic classifiers. All object classifiers described in this manuscript were trained on these perception-evoked patterns.

5.8.2 Characteristic room patterns ("room templates")

To acquire the characteristic neural patterns for rooms ("room templates"), we followed the same procedure that we used for extracting object templates, but here – instead of using post-learning object videos – we used the pre-learning room videos obtained from the first scanning session on day 2 to obtain the characteristic spatial pattern across vertices for every room.

5.8.3 Room reliability

We hypothesized that, in order for a room to serve as an effective retrieval cue for associated memories (i.e., objects paired to rooms), the neural representation for that room must be stable over time and distinct from other room patterns. We captured these properties with a composite measure we called *room reliability*. Crucially, this measure was computed based on data that were collected prior to participants learning the room-object associations. This ensured that our room templates, and therefore our room reliability measure, were not confounded with object information.

To compute room reliability, we obtained the characteristic spatial pattern for each room for each participant, using the procedure outlined above (in the **Characteristic objects patterns** section), but for room videos instead of object videos. Doing this for both pre-learning room video tasks yielded 2 sets of 23 characteristic spatial patterns across vertices (separated in time) for each participant.

We then created a room pattern similarity matrix by correlating the characteristic neural patterns for the rooms from the first pre-learning room video set with the neural patterns for the rooms from the second set. This yielded a 23 x 23 correlation (similarity) matrix for each participant. Because the two pre-learning room videos were separated by a delay, the principal diagonal indicated the similarity of the room representations over time – this was our measure of the stability of the room representations. Similarly, the off-diagonal entries indicated the similarity of one room to another over time, reflecting greater distinctiveness. To create our composite room reliability score for each room, we subtracted the average similarity of the off-diagonal entries (how similar room A is to other rooms over time) from the principal diagonal entry (how similar room A is to itself over time). A large positive difference indicated that a particular room (e.g., room A) was more similar to itself over time than it was to other rooms, indicating its stability and its distinctiveness from other rooms. We did this procedure to obtain a room reliability score for each room of each participant. To quantify significance, for each participant, we averaged reliability across all rooms to get a single difference score per vertex, and performed a 1-sample t-test on these differences against zero before running FDR-correction on the resulting p-values and thresholding at $q < 0.05$.

5.8.4 Object classifier network selection

In order to identify which regions across the brain are involved in the retrieval of object information during guided or free recall, we first needed to identify regions across the brain that could discriminate between objects. To do this, we used two separate phases of the experiment to extract networks that could classify objects during retrieval (when perceptual details of an object were not available) and during perception (when the perceptual details of an object were available). After participants had learned the room-object associations in VR, they were scanned while they watched videos of rooms and asked to recall the objects that were in them (room-video object recall task / post-learning room videos). We used this cued-recall task to identify the retrieval networks (Retrieved Object Classifier Network; ROCN) involved in classifying objects during room videos. Similarly, we identified the networks (Perceived Object

Classifier Network; POCN) involved in perception of objects, by classifying objects during post-learning object videos. Importantly, to avoid circularity in our analyses, all object classifiers (whether those made for ROCN or POCN) were trained with N-1 object perception data using a leave-one-participant-out procedure 25 times, where testing occurred on the left-out participant. The fact that each of the 24 participants in the training dataset had their own set of random room-object pairings ensured that the classifier was able to learn object representations that were not contaminated with room information (by contrast, if we had used a within-participant classification approach, room and object information would have been confounded, since objects were only scanned after they had been paired with a particular room). In other words, because room-object pairings were randomized for every participant, and object evidence for each participant was classified based on object templates derived from the other N-1 participants, any room-related information in the object templates would be unrelated to the room-related information in this left-out participant.

Network selection procedure

In brief, we ran object classifiers on post-learning room videos, where participants had been asked to recall the name of the object paired to the shown room, to identify a network of regions involved in retrieving non-visible object identity. This process involved the following steps: **1)** acquiring the characteristic neural pattern for each object (post-learning object templates); **2)** using a leave-one-participant-out multinomial logistic classifier, trained on the object template patterns for the (N-1) group, to predict object identity in the excluded participant's post-learning room videos (to identify the Retrieved Object Classifier Network) or post-learning object videos (to identify the Perceived Object Classifier Network); and **3)** averaging classifier performance (i.e., accuracy) across all validation searchlights and then selecting the top 50 best classifier searchlights ($\sim$3%). This procedure was done on each searchlight plus the hippocampus ROIs for all participants. Further details are outlined below.

1) Characteristic object patterns (object templates): To extract characteristic neural patterns for objects ("object templates") we used the procedure previously described in the **Characteristic object patterns** methods section.

2) Classifier cross-validation procedure: We applied a leave-one-out cross-validation procedure to predict the left-out participant's object reinstatement at every time point during post-learning room viewing after fitting (i.e., training) a multinomial logistic classifier with the other participants' object pattern templates (i.e., the characteristic spatial patterns estimated from the GLM). More specifically, we shifted the left-out participant's post-learning room video's BOLD timeseries by 4 TRs to approximate the HRF delay, and then trained the classifier with the other participants' object templates before predicting the object class for every timepoint of every room video. To assess the significance of classifier accuracy, we compared the classifier predictions to the correct object class labels and then generated a null distribution of accuracies by shuffling, without replacement, and while preserving their temporal contiguity, the correct object class labels 1000 times; this null distribution was used later to identify searchlights that had above chance accuracy. We did this procedure across all participants such that every participant served as a test participant.

3a) Retrieved Object Classifier Network selection: Post-learning room videos were shown twice to each participant. We ran the leave-one-out cross validation procedure described in the previous section for both runs of the post-learning room viewing separately and then, across all participants and both runs, averaged the classifier accuracy including the corresponding null distributions. We then z-scored the searchlights' (and hippocampus ROIs') performance by comparing the true average accuracies to the average null distribution of accuracies. Afterwards, we extracted the top 50 ROIs with the highest z-scores. This resulted in 50 searchlights (distributed unevenly across hemispheres and excluding hippocampus) corresponding to the searchlights with the top performing classifier performance; these 50 searchlights made up the object retrieval network that we used as an ROI mask in subsequent analyses.

3b) Perceived Object Classifier Network: We applied the same procedure described in the **Retrieved Object Classifier Network selection** section, but instead of classifying non-visible object identity from post-learning room videos, we classified object identity from the post-learning object videos where objects were perceptually visible. In a similar fashion, we extracted the top 50 ROIs by sorting the z-score of accuracies to obtain the network involved in classifying visible objects. Unsurprisingly, the this network was focused on primary visual cortex.

3c) Retrieved Room Classifier Network Selection: We applied a similar procedure described in the **Retrieved Object Classifier Network selection** section, but instead of classifying non-visible object identity from post-learning room videos, we classified room identity from the post-learning object videos where objects *(but not rooms)* were perceptually visible. In a similar fashion, we extracted the top 50 ROIs by sorting the z-score of accuracies to obtain the network involved in classifying room memories. Importantly, the room classifiers were trained on the pre-learning room template patterns for the (N-1) group to predict the recalled room during the held-out participant's post-learning object videos. This ensured that 1) the held-out participant's own room templates were never used for testing, avoiding circularity and 2) the room templates of the group were sourced before any room-object associations were learned, eliminating the potential for these room templates to be contaminated by object information.

5.8.5 Object evidence during guided and free recall

We used the same leave-one-out cross validation procedure described previously to predict object identity during guided and free recalls. As described previously, we shifted each recall timeseries (11 guided recalls and 1 free recall) by 4 TRs to approximate the HRF delay, and used the multinomial classifier to predict object classes at every timepoint for every participant's recalls. Given that that multiclass classifier was trained on all 23 object classes, we obtained a probability distribution across all 23 classes that described the evidence of each class being reinstated at each timepoint. For any specific guided or free recall, we collected the total object evidence across all timepoints a participant verbally recalled that object, regardless of whether the associated room was also recalled. We did not condition our object reinstatement measure on recall of the correctly associated room because we were interested in studying how pre-learning room reliability affects object recall *in general* (as opposed to studying

how reinstatement of a room representation at recall triggers retrieval of the associated object). We then averaged these timepoints across recall runs (guided and free recalls separately). For example, if during the first guided recall a participant verbally recalled the object "teddy bear" in two separate chunks of time for a total of 16 TRs, we collected the classifier probability for "teddy bear" across those 16 TRs and then similarly for every TR "teddy bear" was recalled in all other guided recalls before averaging to get the total "teddy bear" evidence. For a given participant, we did this for each object combining across all 11 guided recalls and, separately, for the participant's single free recall, yielding 23 mean object probabilities for each type of recall task (guided and free recall) for each searchlight.

We wanted to obtain a single value for each object in each participant (separately for guided and free recall), indicating how well that object was reinstated during recall. We did this in two ways: By averaging an object probability across all searchlights that formed part of ROCN or POCN, to obtain an overall ROCN or POCN reinstatement score, respectively, for every object and every participant. We used these as our overall network object reinstatement scores in our subsequent analysis.

5.8.6 Relationship between room reliability and object reinstatement

We hypothesized that rooms with more reliable representations in the pre-learning scans would be associated with higher levels of object reinstatement during self-paced verbal recall. To do this, we ran a searchlight analysis where we correlated the reliability of a room (see **Room reliability** methods section) with the network's evidence for the object paired to that room (see **Object classifier network selection** methods section). We did this for every room-object pair within a participant. For example, for a particular participant, the 23 room reliabilities were correlated with the corresponding 23 object reinstatement probabilities from the retrieval network. Afterwards, we averaged the Fisher-z transformed correlations across participants and recall task types (i.e., guided vs free recall) to generate a single composite correlation map. To test for statistical significance, we ran a nonparametric permutation test in which we randomly shuffled the object labels 1000 times to generate a null distribution of correlations within participants and for both recall types. Significance testing was then performed using the combined null distribution, and resulting p-values were FDR-corrected (see **Fig6A)**. For reference, the results for each recall task type individually are presented in (see **Fig6 - Supp2**).

5.8.7 Benefit of participant-specific room reliability

The analysis shown in **Figure 6A** assesses whether there is a within-participant relationship between room reliability (in a particular searchlight) and object reinstatement. Importantly, there are two possible explanations for this effect (not mutually exclusive). The first is that, within a particular participant, there are idiosyncratic differences in room reliability that predict object reinstatement for that participant – we call this a participant-specific effect. However, there is a second possible explanation: Some rooms may be more reliable than others (averaging across the whole group),

and these generally more reliable rooms may support better object reinstatement on average (e.g., the chess room might consistently be better represented across people and support better object recall) – we call this a group-wise effect. Both kinds of effect are important, but they have different connotations: If the relationship between room reliability and object reinstatement is driven by idiosyncratic (participant-specific) factors, then there is predictive value in doing a "personalized audit" of the person's memory palace by scanning them; but if there is only a group-wise relationship, there is no need to collect scanning data from a new person, so long as you already have data on room reliability from the rest of the group. To assess whether the observed within-subjects relationship between room reliability and object reinstatement has a participant-specific component, we compared the predictive performance of an ordinary least squares regression derived from a participant's own room reliability values with one based on other individuals' reliability values. Specifically, for the participant-specific model (as in **Fig 6A**), we calculated the coefficient of determination (R^2) of a model where the participant's object reinstatement probabilities were predicted by their own room reliability values. For the group-wise model, we iteratively predicted that participant's reinstatement probabilities from every single other participant and averaged the resulting N-1 R^2 values. We then ran a model comparison test where we took the difference between the R^2 of the participant-specific model and the average R^2 from the other-participant prediction models. A significant positive difference in this analysis indicates that the participant-specific model explains the variability in object evidence better than other individuals (and thus the observed results can not be entirely due to the group-wise effect). To test for statistical significance, we ran a nonparametric permutation test where the object labels were randomly shuffled 1000 times to generate a null distribution of model performance for each model. To generate a single composite map summarizing model performance across recall tasks, we averaged R^2 values across guided and free recalls for each model separately, computed the difference in R^2, and tested for significance using a nonparametric permutation test on the combined null distribution of R^2 differences, followed by FDR-correction on the resulting p-values. For completeness, separate results for guided and free recall are provided in **Fig6 - Supp2**, while the main results reflect the combined composite analysis (see **Fig6**).

5.8.8 Partial correlation analysis controlling for room reinstatement

To test whether room reliability predicted subsequent object reinstatement when controlling for room reinstatement at recall, we conducted a partial correlation analysis. Specifically, we asked whether the correlation between room reliability and object reinstatement (**Fig6A**) remained significant after regressing room reinstatement at recall out of both of these other variables.

To do this, we first constructed a Retrieved Room Classifier Network (RRCN). As described in the **Network Selection Procedure** Methods section, we followed a similar approach to identify ROCN and POCN. In brief, we used a leave-one-participant-out cross-validation procedure in which we classified room recall during the held-out participant's perception of object videos. The top 50 best performing

searchlights were used to define the RRCN, which was then used as a mask to extract room reinstatement evidence for our partial correlation analysis.

In this analysis, we wanted to control for room reinstatement that occurred on timepoints when participants verbally recounted room details *and* on timepoints when they verbally described the objects that were paired to a particular room – in principle, room reinstatement during either set of timepoints could be acting to scaffold object retrieval. To this end, we computed two separate room reinstatement scores within the RRCN:

RRCN-room-recall: Room evidence extracted with the RRCN mask during timepoints in which participants were speaking about a room during free and guided recall.

RRCN-object-recall: Room evidence extracted within the RRCN mask during timepoints in which participants were speaking about the object that had been associated with a given room during guided and free recall.

To isolate the unique relationship between room reliability and object reinstatement, we regressed out both RRCN measures from each variable. Specifically, we fit a linear model with ROCN object reinstatement as the dependent variable and both RRCN-room-recall and RRCN-object-recall as predictors. The ROCN residuals from this model represented object reinstatement variance unexplained by room reinstatement. Similarly, we fit a second linear model with room reliability as the dependent variable and the same two RRCN measures as predictors. The room reliability residuals from this model represented room reliability variance unexplained by room reinstatement. Finally, we computed a Pearson correlation between these two residuals. To test for significance, we ran a nonparametric permutation test in which we shuffled the ROCN residuals and recomputed the correlation 1000 times to generate a null distribution of correlation values before running FDR correction for $q < 0.05$.

To identify regions where the relationship between room reliability and object reinstatement had a significant positive or negative change after controlling for room reinstatement, we ran a contrast in which the correlation values of the partial correlation were subtracted from the correlation values of our original model. To test for this difference, we computed a composite score of each correlation by averaging the results of each searchlight across recall task types (i.e., guided and free recalls) and participants. Next, we computed the difference between the results of our original model and the partial correlation as well as on their permutations to get a null distribution of differences. To test for significance, we ran a nonparametric permutation test where we compared the true differences from the null-distribution of differences and FDR corrected for $q < 0.05$.

5.8.9 Relationship between room reliability and room features

Do properties of a room contribute to the reliability of their representation? We sought to identify whether physical or graph theoretical features of a room contributed to the reliability of their representation. To do this we used the 3D Unity model of the environment to compute a list of physical features such as total room volume, total volume occupied by background objects, the proportion of occupied volume and total room volume, area, object count, number of corners, and whether the room has a

window (i.e., a view to the outside), and used the room adjacency matrix to compute graph-theoretical features such as degree, betweenness, closeness, eigenvector, and pagerank. We then selected six features (degree, ratio of occupied volume, background object count, floor area, number of corners, and "has window") that were the least collinear and provided conceptually non-overlapping properties (e.g., betweenness and degree are collinear). We z-scored each feature (except the binary "has window") and then ran a searchlight analysis where we regressed room reliability on each of the six z-scored features for every participant. To test for statistical significance of each of the resulting beta coefficients, we ran a nonparametric permutation test where room reliability was shuffled 1000 times within participants before regressing again on the features to generate a null distribution of beta coefficients. We then averaged across participants before running FDR-correction on the resulting z-values and thresholding at $q < 0.001$.

5.9 Data and code

All data are openly available for download here:
https://openneuro.org/datasets/ds005704
Scripts used for analysis also available here:
https://github.com/rmasiso/MemoryPalaceReliability

6 Acknowledgements

We would like to thank Neggin Keshavarzian, Mai Nguyen, Hanna Hillman, Sam Nastase, Sam Zorowitz, Thea Zalaback, Silvy Collin, James Antony and everyone else for buddying during the long fMRI sessions, Norbert Cruz-Lebrón, Jamal Williams, the Norman lab and Baldassano lab members for insightful comments and feedback. Sam Perrin, the VRTK, and Unity community for valuable insight on optimizing VR gameplay. Jamal, Jake Reske, Nathan Prillaman, Javier Masís and Burne Holiday (Javier, Joey Edelmann, Cory Furlong, and Nathan Tyrell) for providing beautiful music to accompany the rooms. We would also like to thank Janice Chen and Chris Honey for their guidance and support. This work was supported by a Multi-University Research Initiative Grant (ONR/DoDN00014-17-1-2961) to KAN and NINDS D-SPAN F99NS120644-01 and T32MH065214 to RMO.

References

Avants, B.B., Epstein, C.L., Grossman, M., Gee, J.C.: Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. Medical image analysis **12**(1), 26–41 (2008)

Arnold, K.M., McDermott, K.B., Szpunar, K.K.: Imagining the near and far future: The role of location familiarity. Memory & Cognition **39**, 954–967 (2011)

Baldassano, C., Chen, J., Zadbood, A., Pillow, J.W., Hasson, U., Norman, K.A.: Discovering event structure in continuous narrative perception and memory. Neuron **95**(3), 709–721 (2017) https://doi.org/10.1016/j.neuron.2017.06.041

Baldassano, C., Hasson, U., Norman, K.A.: Representation of real-world event schemas during narrative perception. Journal of Neuroscience **38**(45), 9689–9699 (2018) https://doi.org/10.1523/JNEUROSCI.0251-18.2018

Bramao, I., Karlsson, A., Johansson, M.: Mental reinstatement of encoding context improves episodic remembering. Cortex **94**, 15–26 (2017)

Brunec, I.K., Moscovitch, M., Barense, M.D.: Boundaries shape cognitive representations of spaces and events. Trends in Cognitive Sciences **22**(7), 637–650 (2018) https://doi.org/10.1016/j.tics.2018.03.013

Burgess, N., Maguire, E.A., Spiers, H.J., O'Keefe, J.: A temporoparietal and prefrontal network for retrieving the spatial context of lifelike events. Neuroimage **14**(2), 439–453 (2001)

Brodt, S., Pöhlchen, D., Flanagin, V.L., Glasauer, S., Gais, S., Schönauer, M.: Rapid and independent memory formation in the parietal cortex. Proceedings of the National Academy of Sciences **113**(46), 13251–13256 (2016)

Bower, G.H., Reitman, J.S.: Mnemonic elaboration in multilist learning. Journal of verbal learning and verbal behavior **11**(4), 478–485 (1972)

Behzadi, Y., Restom, K., Liau, J., Liu, T.T.: A component based noise correction method (compcor) for bold and perfusion based fmri. Neuroimage **37**(1), 90–101 (2007)

Chang, M., Hong, B., Savel, K., Du, J., Meade, M.E., Martin, C.B., Barense, M.D.: Spatial context scaffolds long-term episodic richness of weaker real-world autobiographical memories in both older and younger adults. Memory **32**(4), 431–448 (2024)

Cox, R.W.: AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. Computers and Biomedical Research **29**(3), 162–173 (1996) https://doi.org/10.1006/cbmr.1996.0014

Cooper, R.A., Ritchey, M.: Cortico-hippocampal network connections support the multidimensional quality of episodic memory. Elife **8**, 45591 (2019)

Dale, A.M., Fischl, B., Sereno, M.I.: Cortical surface-based analysis: I. segmentation and surface reconstruction. Neuroimage **9**(2), 179–194 (1999)

Dordevic, M., Hoelzer, S., Russo, A., García Alanis, J.C., Müller, N.G.: The role of the precuneus in human spatial updating in a real environment setting—a ctbs study.

Life **12**(8), 1239 (2022)

Dresler, M., Shirer, W.R., Konrad, B.N., Müller, N.C., Wagner, I.C., Fernández, G., Czisch, M., Greicius, M.D.: Mnemonic training reshapes brain networks to support superior memory. Neuron **93**(5), 1227–1235 (2017)

De Soares, A., Kim, T., Mugisho, F., Zhu, E., Lin, A., Zheng, C., Baldassano, C.: Top-down attention shifts behavioral and neural event boundaries in narratives with overlapping event scripts. bioRxiv, 2023–08 (2023)

Epstein, R.A., Baker, C.I.: Scene perception in the human brain. Annual review of vision science **5**(1), 373–397 (2019)

Epstein, R.A., Patai, E.Z., Julian, J.B., Spiers, H.J.: The cognitive map in humans: spatial navigation and beyond. Nature neuroscience **20**(11), 1504–1513 (2017)

Essoe, J.K.-Y., Reggente, N., Ohno, A.A., Baek, Y.H., Dell'Italia, J., Rissman, J.: Enhancing learning and retention with distinctive virtual reality environments and mental context reinstatement. npj Science of Learning **7**(1), 31 (2022)

Fonov, V.S., Evans, A.C., McKinstry, R.C., Almli, C.R., Collins, D.: Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. NeuroImage (47), 102 (2009)

Flanagin, V.L., Klinkowski, S., Brodt, S., Graetsch, M., Roselli, C., Glasauer, S., Gais, S.: The precuneus as a central node in declarative memory retrieval. Cerebral Cortex (2023)

Fellner, M.-C., Volberg, G., Wimber, M., Goldhacker, M., Greenlee, M.W., Hanslmayr, S.: Spatial mnemonic encoding: theta power decreases and medial temporal lobe bold increases co-occur during the usage of the method of loci. eneuro **3**(6) (2016)

Frings, L., Wagner, K., Quiske, A., Schwarzwald, R., Spreer, J., Halsband, U., Schulze-Bonhage, A.: Precuneus is involved in allocentric spatial location encoding and recognition. Experimental Brain Research **173**, 661–672 (2006)

Godden, D.R., Baddeley, A.D.: Context-dependent memory in two natural environments: on land and underwater. British Journal of Psychology **66**(3), 325–331 (1975) https://doi.org/10.1111/j.2044-8295.1975.tb01468.x

Gorgolewski, K., Burns, C.D., Madison, C., Clark, D., Halchenko, Y.O., Waskom, M.L., Ghosh, S.S.: Nipype: a flexible, lightweight and extensible neuroimaging data processing framework in python. Frontiers in neuroinformatics, 13 (2011)

Gorgolewski, K.J., Esteban, O., Ellis, D., Notter, M., Ziegler, E., Johnson, H., Hamalainen, C., Yvernault, B., Burns, C., Manhães-Savio, A., et al.: Nipype: a flexible, lightweight and extensible neuroimaging data processing framework in python.

0.13. 1. Zenodo (2017)

Greve, D.N., Fischl, B.: Accurate and robust brain image alignment using boundary-based registration. Neuroimage **48**(1), 63–72 (2009)

Gilboa, A., Marlatte, H.: Neurobiology of schemas and schema-mediated memory. Trends in Cognitive Sciences **21**(8), 618–631 (2017) https://doi.org/10.1016/j.tics.2017.04.013

Ghaem, O., Mellet, E., Crivello, F., Tzourio, N., Mazoyer, B., Berthoz, A., Denis, M.: Mental navigation along memorized routes activates the hippocampus, precuneus, and insula. Neuroreport **8**(3), 739–744 (1997)

Graesser, A.C., Nakamura, G.V.: The impact of a schema on comprehension and memory. Psychology of Learning and Motivation - Advances in Research and Theory **16**(C), 59–109 (1982) https://doi.org/10.1016/S0079-7421(08)60547-2

Gurguryan, L., Sheldon, S.: Retrieval orientation alters neural activity during autobiographical memory recollection. Neuroimage **199**, 534–544 (2019)

Guo, D., Yang, J.: Interplay of the long axis of the hippocampus and ventromedial prefrontal cortex in schema-related memory retrieval. Hippocampus **30**(3), 263–277 (2020) https://doi.org/10.1002/hipo.23154

Hasson, U., Chen, J., Honey, C.J.: Hierarchical process memory: memory as an integral component of information processing. Trends in Cognitive Sciences **19**(6), 304–313 (2015) https://doi.org/10.1016/j.tics.2015.04.006 . Publisher: Elsevier. Accessed 2020-03-01

Hirshhorn, M., Grady, C., Rosenbaum, R.S., Winocur, G., Moscovitch, M.: Brain regions involved in the retrieval of spatial and episodic details associated with a familiar environment: an fmri study. Neuropsychologia **50**(13), 3094–3106 (2012)

Hebscher, M., Levine, B., Gilboa, A.: The precuneus and hippocampus contribute to individual differences in the unfolding of spatial representations during episodic autobiographical memory. Neuropsychologia **110**, 123–133 (2018)

Hassabis, D., Maguire, E.A.: Deconstructing episodic memory with construction. Trends in cognitive sciences **11**(7), 299–306 (2007)

Huang, J., Manglik, A., Dutra, N., Tarder-Stoll, H., Chamberlain, T., Ajemian, R., Zhang, Q., Norman, K.A., Baldassano, C.: Binding items to contexts through conjunctive neural representations with the method of loci. bioRxiv, 2024–12 (2024)

Hartley, T., Maguire, E.A., Spiers, H.J., Burgess, N.: The well-worn route and the path less traveled: distinct neural bases of route following and wayfinding in humans. Neuron **37**(5), 877–888 (2003)

Hayes, S.M., Ryan, L., Schnyer, D.M., Nadel, L.: An fmri study of episodic memory: retrieval of object, spatial, and temporal information. Behavioral neuroscience **118**(5), 885 (2004)

Herweg, N.A., Sharan, A.D., Sperling, M.R., Brandt, A., Schulze-Bonhage, A., Kahana, M.J.: Reactivated spatial context guides episodic recall. Journal of Neuroscience **40**(10), 2119–2128 (2020)

Huntenburg, J.M.: Evaluating nonlinear coregistration of bold epi and t1w images. PhD thesis, Freie Universität Berlin (2014)

Hasson, U., Yang, E., Vallines, I., Heeger, D.J., Rubin, N.: A hierarchy of temporal receptive windows in human cortex. The Journal of Neuroscience **28**(10), 2539–2550 (2008)

Ino, T., Inoue, Y., Kage, M., Hirose, S., Kimura, T., Fukuyama, H.: Mental navigation in humans is processed in the anterior bank of the parieto-occipital sulcus. Neuroscience letters **322**(3), 182–186 (2002)

Jenkinson, M., Bannister, P., Brady, M., Smith, S.: Improved optimization for the robust and accurate linear registration and motion correction of brain images. Neuroimage **17**(2), 825–841 (2002)

Konkle, T., Brady, T.F., Alvarez, G.A., Oliva, A.: Conceptual distinctiveness supports detailed visual long-term memory for real-world objects. Journal of experimental Psychology: general **139**(3), 558 (2010)

Klein, A., Ghosh, S.S., Bao, F.S., Giard, J., Häme, Y., Stavsky, E., Lee, N., Rossa, B., Reuter, M., Chaibub Neto, E., *et al.*: Mindboggling morphometry of human brains. PLoS computational biology **13**(2), 1005350 (2017)

Kondo, Y., Suzuki, M., Mugikura, S., Abe, N., Takahashi, S., Iijima, T., Fujii, T.: Changes in brain activation associated with use of a memory strategy: a functional mri study. Neuroimage **24**(4), 1154–1163 (2005)

Lerner, Y., Honey, C.J., Silbert, L.J., Hasson, U.: Topographic mapping of a hierarchy of temporal receptive windows using a narrated story. The Journal of Neuroscience **31**(8), 2906–2915 (2011)

Legge, E.L., Madan, C.R., Ng, E.T., Caplan, J.B.: Building a memory palace in minutes: Equivalent memory performance using virtual versus conventional environments with the method of loci. Acta psychologica **141**(3), 380–390 (2012)

Lohnas, L.J., Polyn, S.M., Kahana, M.J.: Contextual variability in free recall. Journal of memory and language **64**(3), 249–255 (2011)

Liu, C., Ye, Z., Chen, C., Axmacher, N., Xue, G.: Hippocampal representations of event

structure and temporal context during episodic temporal order memory. Cerebral Cortex **32**(7), 1520–1534 (2022)

Mallow, J., Bernarding, J., Luchtmann, M., Bethmann, A., Brechmann, A.: Superior memorizers employ different neural networks for encoding and recall. Frontiers in Systems Neuroscience, 128 (2015)

Moè, A., De Beni, R.: Stressing the efficacy of the loci method: oral presentation and the subject-generation of the loci pathway with expository passages. Applied Cognitive Psychology **19**(1), 95–106 (2005)

Melton, A.W.: The situation with respect to the spacing of repetitions and memory. Journal of Verbal Learning and Verbal Behavior **9**(5), 596–606 (1970)

Maguire, E.A., Frackowiak, R.S., Frith, C.D.: Recalling routes around london: activation of the right hippocampus in taxi drivers. Journal of neuroscience **17**(18), 7103–7110 (1997)

Miller, J.F., Lazarus, E.M., Polyn, S.M., Kahana, M.J.: Spatial clustering during memory search. Journal of Experimental Psychology: Learning, Memory, and Cognition **39**(3), 773–781 (2013)

McClelland, J.L., McNaughton, B.L., O'Reilly, R.C.: Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. Psychological review **102**(3), 419 (1995)

Masís-Obando, R., Norman, K.A., Baldassano, C.: Schema representations in distinct brain networks support narrative memory during encoding and retrieval. Elife **11**, 70445 (2022)

Maguire, E.A., Valentine, E.R., Wilding, J.M., Kapur, N.: Routes to remembering: the brains behind superior memory. Nature neuroscience **6**(1), 90–95 (2003)

Nyberg, L., Sandblom, J., Jones, S., Neely, A.S., Petersson, K.M., Ingvar, M., Bäckman, L.: Neural correlates of training-related memory improvement in adulthood and aging. Proceedings of the National Academy of Sciences **100**(23), 13728–13733 (2003)

Preston, A.R., Eichenbaum, H.: Interplay of hippocampus and prefrontal cortex in memory. Current Biology **23**(17), 764–773 (2013) https://doi.org/10.1016/j.cub.2013.05.041

Peirce, J.W.: PsychoPy-Psychophysics software in Python. Journal of Neuroscience Methods **162**(1-2), 8–13 (2007) https://doi.org/10.1016/j.jneumeth.2006.11.017

Poppenk, J., Evensmoen, H.R., Moscovitch, M., Nadel, L.: Long-axis specialization

of the human hippocampus. Trends in Cognitive Sciences **17**(5), 230–240 (2013) https://doi.org/10.1016/j.tics.2013.03.005

Power, J.D., Mitra, A., Laumann, T.O., Snyder, A.Z., Schlaggar, B.L., Petersen, S.E.: Methods to detect, characterize, and remove motion artifact in resting state fmri. Neuroimage **84**, 320–341 (2014)

Ranganath, C.: A unified framework for the functional organization of the medial temporal lobes and the phenomenology of episodic memory. Hippocampus **20**(11), 1263–1290 (2010)

Robin, J., Buchsbaum, B.R., Moscovitch, M.: The primacy of spatial context in the neural representation of events. Journal of Neuroscience **38**(11), 2755–2765 (2018) https://doi.org/10.1523/JNEUROSCI.1638-17.2018 . Accessed 2020-01-09

Reggente, N., Essoe, J.K.-Y., Aghajan, Z.M., Tavakoli, A.V., McGuire, J.F., Suthana, N.A., Rissman, J.: Enhancing the ecological validity of fMRI memory research using virtual reality. Frontiers in Neuroscience **12**, 408 (2018) https://doi.org/10.3389/fnins.2018.00408 . Accessed 2020-03-12

Reggente, N., Essoe, J.K., Baek, H.Y., Rissman, J.: The method of loci in virtual reality: explicit binding of objects to spatial contexts enhances subsequent memory recall. Journal of Cognitive Enhancement **4**, 12–30 (2020)

Robin, J., Garzon, L., Moscovitch, M.: Spontaneous memory retrieval varies based on familiarity with a spatial context. Cognition **190**, 81–92 (2019)

Robin, J., Moscovitch, M.: The effects of spatial contextual familiarity on remembered scenes, episodic memories, and imagined future events. Journal of Experimental Psychology: Learning, Memory, and Cognition **40**(2), 459 (2014)

Robin, J., Moscovitch, M.: Familiar real-world spatial cues provide memory benefits in older and younger adults. Psychology and Aging **32**(3), 210 (2017)

Roediger, H.L.: The effectiveness of four mnemonics in ordering recall. Journal of Experimental Psychology: Human Learning and Memory **6**(5), 558 (1980)

Reagh, Z.M., Ranganath, C.: Flexible reuse of cortico-hippocampal representations during encoding and recall of naturalistic events. Nature Communications **14**(1), 1279 (2023)

Robin, J., Wynn, J., Moscovitch, M.: The spatial scaffold: The effects of spatial context on memory for events. Journal of Experimental Psychology: Learning, Memory, and Cognition **42**(2), 308 (2016)

Rosenbaum, R.S., Ziegler, M., Winocur, G., Grady, C.L., Moscovitch, M.: "i have often walked down this street before": fmri studies on the hippocampus and other

structures during mental navigation of an old environment. Hippocampus **14**(7), 826–835 (2004)

Schank, R.C., Abelson, R.P.: Scripts, Plans, Goals, and Understanding: An Inquiry Into Human Knowledge Structures. Artificial Intelligence Series. Lawrence Erlbaum Associates, Hillsdale, NJ (1977). https://books.google.com/books?id=YZ99AAAAMAAJ

Sievers, C., Bird, C.M., Renoult, L.: Predicting memory formation over multiple study episodes. Learning & Memory **26**(12), 465–472 (2019)

Sheldon, S., Chu, S.: What versus where: Investigating how autobiographical memory retrieval differs when accessed with thematic versus spatial information. The Quarterly Journal of Experimental Psychology **70**(9), 1909–1921 (2017)

Sheldon, S., Cool, K., El-Asmar, N.: The processes involved in mentally constructing event-and scene-based autobiographical representations. Journal of Cognitive Psychology **31**(3), 261–275 (2019)

Sheldon, S., Gurguryan, L., Madore, K.P., Schacter, D.L.: Constructing autobiographical events within a spatial or temporal context: a comparison of two targeted episodic induction techniques. Memory **27**(7), 881–893 (2019)

Shin, Yeon Soon and Masís-Obando, Rolando, Keshavarzian, N., Dáve, R., Norman, K.A.: Context-dependent memory effects in two immersive virtual reality environments: On mars and underwater. Psychonomic Bulletin & Review **28**(2), 574–582 (2021)

Smith, S.M.: Remembering in and out of context. Journal of Experimental Psychology: Human Learning and Memory **5**(5), 460 (1979)

Spreng, R.N., Mar, R.A., Kim, A.S.: The common neural basis of autobiographical memory, prospection, navigation, theory of mind, and the default mode: a quantitative meta-analysis. Journal of cognitive neuroscience **21**(3), 489–510 (2009)

Sommer, V.R., Sander, M.C.: Contributions of representational distinctiveness and stability to memory performance and age differences. Aging, Neuropsychology, and Cognition **29**(3), 443–462 (2022)

Szpunar, K.K., St. Jacques, P.L., Robbins, C.A., Wig, G.S., Schacter, D.L.: Repetition-related reductions in neural activity reveal component processes of mental simulation. Social Cognitive and Affective Neuroscience **9**(5), 712–722 (2014)

Salan, J., Smith, D.E., Shafer, E.S., Diana, R.A.: Variation in encoding context benefits item recognition. Memory & Cognition, 1–21 (2024)

Smith, S.M., Vela, E.: Environmental context-dependent memory: a review and meta-analysis. Psychonomic Bulletin & Review **8**(2), 203–220 (2001)

Schott, B.H., Wüstenberg, T., Lücke, E., Pohl, I.-M., Richter, A., Seidenbecher, C.I., Pollmann, S., Kizilirmak, J.M., Richardson-Klavehn, A.: Gradual acquisition of visuospatial associative memory representations via the dorsal precuneus. Human Brain Mapping **40**(5), 1554–1570 (2018)

Tustison, N.J., Avants, B.B., Cook, P.A., Zheng, Y., Egan, A., Yushkevich, P.A., Gee, J.C.: N4itk: improved n3 bias correction. IEEE transactions on medical imaging **29**(6), 1310–1320 (2010)

Treiber, J.M., White, N.S., Steed, T.C., Bartsch, H., Holland, D., Farid, N., McDonald, C.R., Carter, B.S., Dale, A.M., Chen, C.C.: Characterization and correction of geometric distortions in 814 diffusion weighted images. PloS one **11**(3), 0152472 (2016)

Wagner, I.C., Konrad, B.N., Schuster, P., Weisig, S., Repantis, D., Ohla, K., Kühn, S., Fernández, G., Steiger, A., Lamm, C., *et al.*: Durable memories and efficient neural coding through mnemonic training using the method of loci. Science advances **7**(10), 7606 (2021)

Wang, S., Peterson, D.J., Gatenby, J.C., Li, W., Grabowski, T.J., Madhyastha, T.M.: Evaluation of field map and nonlinear registration methods for correction of susceptibility artifacts in diffusion mri. Frontiers in neuroinformatics **11**, 17 (2017)

Xue, G., Dong, Q., Chen, C., Lu, Z., Mumford, J.A., Poldrack, R.A.: Greater neural pattern similarity across repetitions is associated with better memory. Science **330**(6000), 97–101 (2010) https://doi.org/10.1126/science.1193125

Yonelinas, A.P., Ranganath, C., Ekstrom, A.D., Wiltgen, B.J.: A contextual binding theory of episodic memory: systems consolidation reconsidered. Nature Reviews Neuroscience **20**(6), 364–375 (2019)

Zhang, Y., Brady, M., Smith, S.: Segmentation of brain mr images through a hidden markov random field model and the expectation-maximization algorithm. IEEE transactions on medical imaging **20**(1), 45–57 (2001)

7 Supplementary Information

Supplementary Table 1 Virtual memory palace room music

Room id	Room Name	Track Name	Track Artist
1	Antiques Room	Rio De Colores	Strunz & Farah
2	TV Room	Birds	Rolando Masís-Obando
3	Candy Room	Liar Liar	The Castaways
4	Classroom	Scatting at the 18th Grammy Awards, Feb. 1976	Ella Fitzgerald & Mel Tormé
5	Tool Room	brazil loops 1	motion_correct
6	Bedroom	Rero	Burne Holiday
7	Floating Islands Room	Brooks Was Here (800% slower) (Rolando Masís-Obando StretchMix)	Thomas Newman
8	Planet Room	Electric à la Mass Effect	Rolando Masís-Obando
9	Computer Store Room	Inmersión	Éditus 360
10	Painting Room	Juana y Mi Hermana	Sonámbulo Psicotropical
11	Storage Boxes Room	Aragorn's Speech at the Black Gate	from Lord of the Rings
12	Chess Room	Prelude in G Minor (Op. 23 No. 5)	Sergei Rachmaninoff
13	Empty Room	Wrong Side	Strapping Young Lad
14	Cat Portraits Room	Symphony No.8 "Unfinished" D 759	Franz Schubert
15	Ruins Room	Memory (Extended)	Nathan Prillaman
16	Clocks Room	Birds Stretched	Rolando Masís-Obando
17	Crystals Room	colorthought - crash _JR 11.4.16	Jakob Reske
18	Colorful Wall Room	Zaggazagga	Rolando Masís-Obando
19	Altar Room	Throatsinging	Rolando Masís-Obando
20	Applecreates Room	Tron Overture (from the Motion Picture TRON: Legacy)	Daft Punk
21	Birthday Party Room	Birthday Horror	Rolando Masís-Obando & the Ken Norman Lab
22	Firepit Room	fire crackles	N/A
23	Human Portraits Room	What Way?	Javier Masís

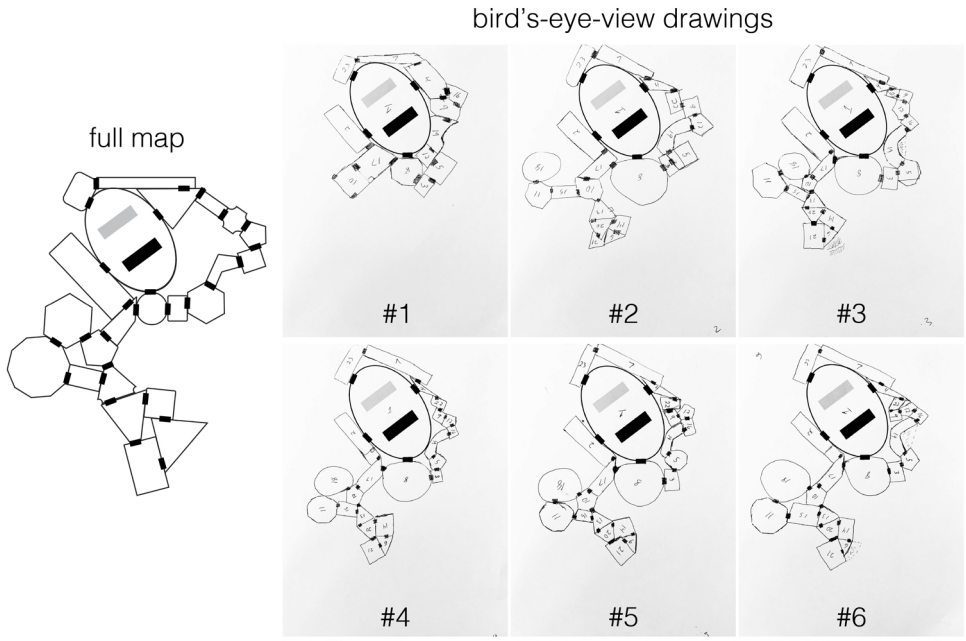

Figure 1–Figure supplement 1. Example bird's-eye-view drawings from a single selected participant. Each participant was asked to draw multiple bird's-eye-view maps of the environment across pre- and post-learning sessions to monitor training progress. Each participant was given two sheets of paper, one for the map they had to draw with a central room drawn in, and another with a legend containing the room name and pre-determined room number to facilitate map drawing and room labeling.

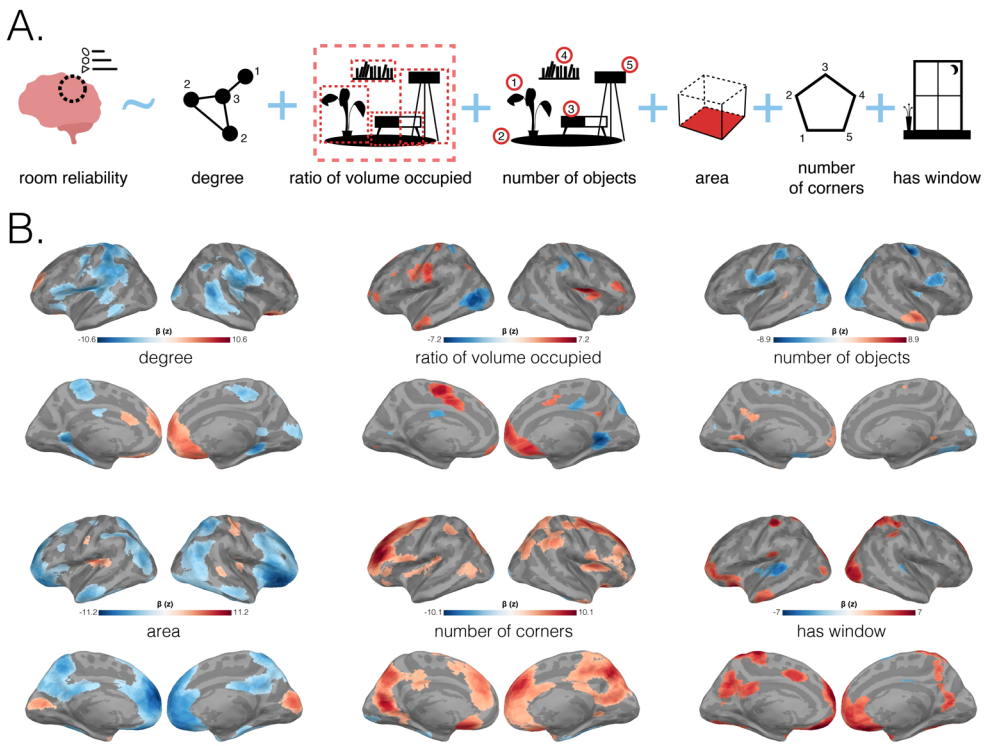

Figure 2–Figure supplement 1. Relationship between room reliability and room features. (A.) Regression schematic predicting room reliability with room features. Six different room features were chosen to predict room reliability. From left to right: "degree" (how many rooms are connected to room of interest), "ratio of volume occupied" (the proportion of volume occupied by objects inside a room), "number of objects" (manual count of every object inside a room), "area" (area covered by room floor), "number of corners" (sum of wall corners in room), and "has window" (binary, indicating whether this room has a view to the outside). **(B.)** Significant room feature regression coefficients. In a searchlight analysis, reliability for a room (in that searchlight) was predicted by six different room features for each participant. Statistical significance for the resulting beta coefficients was determined by a non-parametric permutation test and FDR-corrected for $q < 0.001$. All six surface maps are colored based on the magnitude of significant z-scored coefficients, with blue showing negative and red showing positive relationships, respectively.

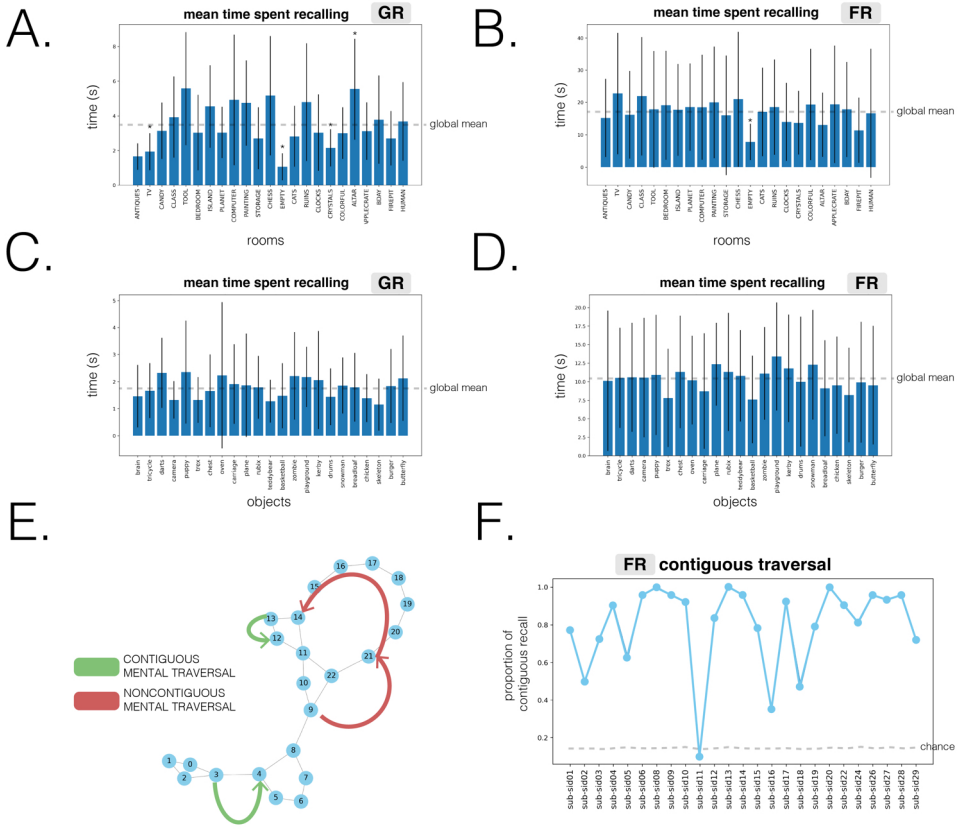

Figure 3–Figure supplement 1. Additional behavioral measures. (A–D) Average time participants spent during recall tasks: **(A.)** Recalling rooms during guided recalls. **(B.)** Recalling rooms during free recalls. **(C.)** Recalling objects during guided recalls. **(D.)** Recalling objects during free recalls. For **A-D**, bar charts describe mean recall time across participants, with error bars indicating standard deviation, and dashed gray lines indicating the global mean. Stars above bar plots indicate a significant difference in time from the global mean, for Bonferroni corrected $p < 0.05$. **(E.)** Contiguous vs noncontiguous mental traversal. When recalling rooms explored in VR, participants either recalled spatially adjacent (contiguous) rooms or jumped between rooms separated by more than one degree (noncontiguous). **(F.)** Contiguity in free recall. We measured the proportion of contiguous room transitions during free recall. Across participants, spatially adjacent rooms were recalled more often than expected by chance ($t(24) = 14.19, p < 0.001$) suggesting an unprompted bias towards contiguous mental traversal.

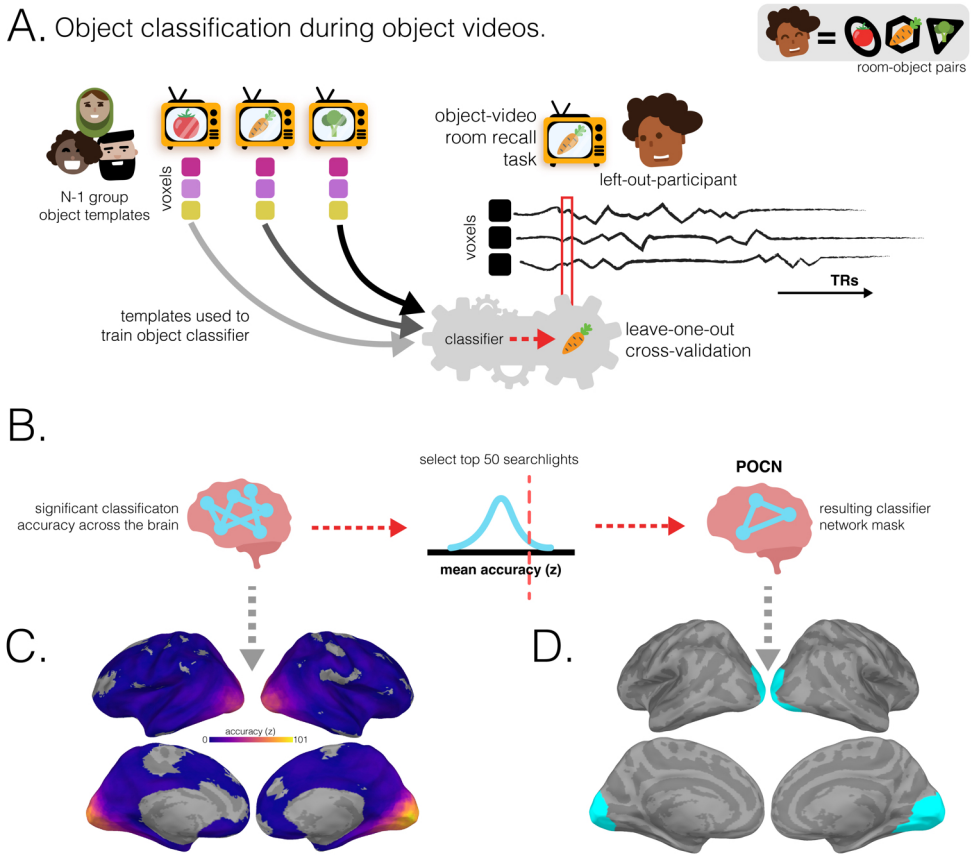

Figure 4–Figure supplement 1. Perceived object classifier network (POCN) methodology and surface maps. (A.) As for the ROCN, the characteristic object patterns of the N-1 group were used to train a multinomial logistic classifier. Here, this classifier was applied to timepoints when the left-out participant was viewing objects during the post-learning object videos, e.g. viewing a video of a carrot in this example. We then measured the fraction of timepoints during the object video that were classified as activating the carrot representation. **(B.)** After object classification was performed for both post-learning object video runs for each participant, average classification accuracies across participants for each searchlight were then averaged across both runs and then z-scored relative to a null distribution. The top 50 searchlights were then selected to form the POCN. **(C.)** Average object classification accuracy during object encoding (thresholded to show searchlights with above-chance accuracy). **(D.)** Perceived Object Classifier Network. The surface map shows the top 50 best object classifying searchlights during both post-learning object video tasks across participants.

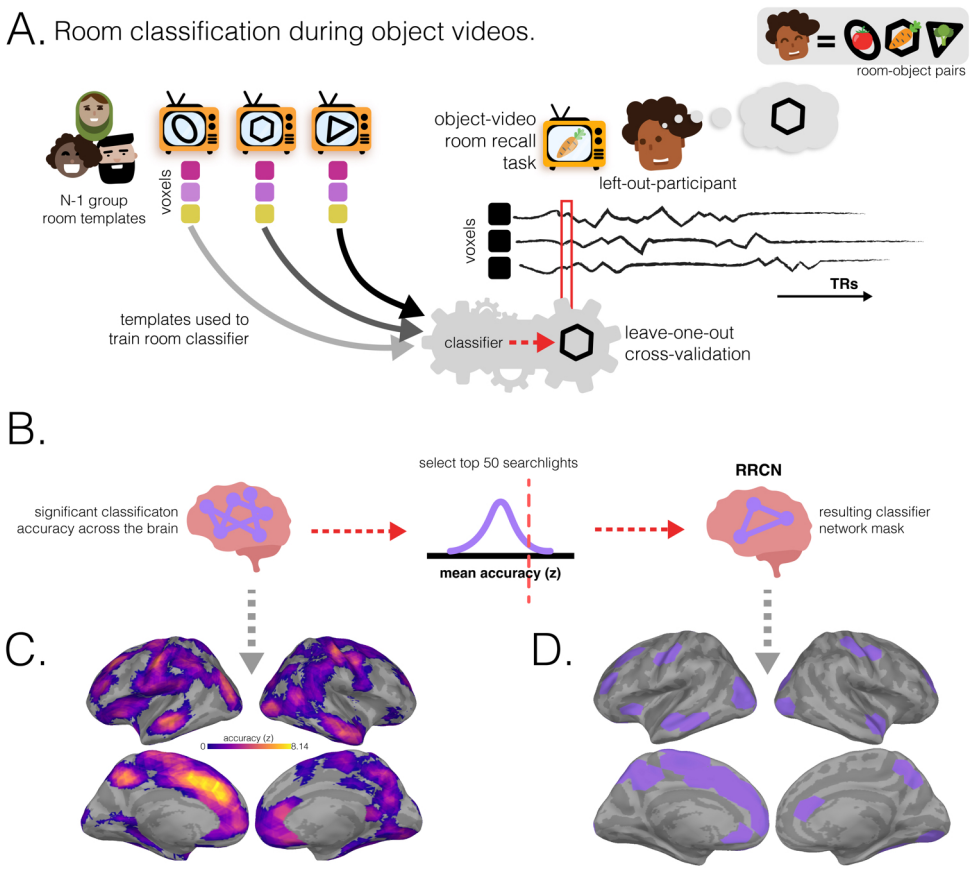

Figure 4–Figure supplement 2. Retrieved room classifier network (RRCN) methodology and surface maps. (A.) The characteristic room patterns of the N-1 group – evoked during a separate phase of the study in which participants viewed room videos before learning room-object associations – were used to train a multinomial logistic classifier. This classifier was then applied to each timepoint on the left-out participant's object-video room recall data. In the pictured example, the left-out participant, Fernando, is recalling the hexagon room that was paired with the carrot object currently being presented. The room classifier, trained on patterns evoked when other participants viewed the rooms, was applied to each timepoint of Fernando's carrot viewing. We then measured the fraction of timepoints during the object video that were classified as activating the hexagon representation. **(B.)** After room classification was performed for both post-learning object video runs for each participant, average classification accuracies across participants for each searchlight were then averaged across both runs and then z-scored relative to a null distribution. The top 50 searchlights were then selected to form the RRCN. **(C.)** Average object classification accuracy during object encoding (thresholded to show searchlights with above-chance accuracy). **(D.)** Retrieved Room Classifier Network. The surface map shows the top 50 best room classifying searchlights during both post-learning object video tasks across participants.

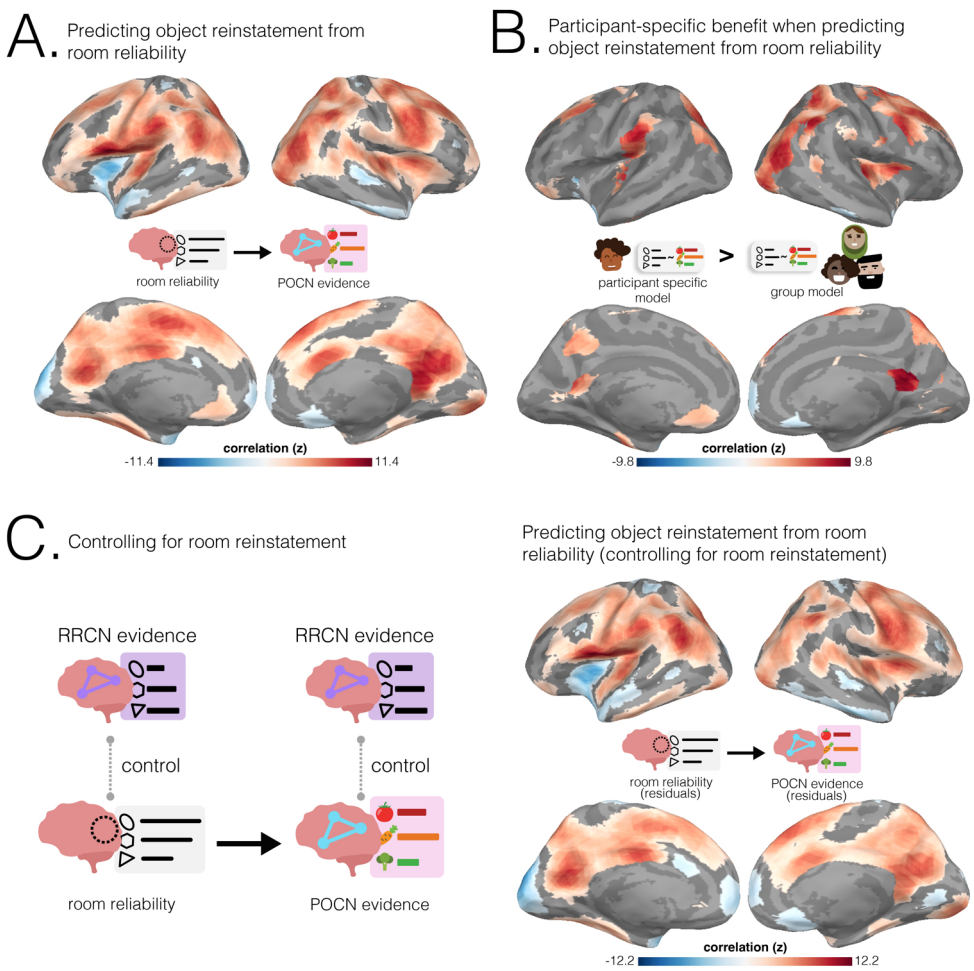

Figure 6–Figure supplement 1.
Predicting POCN object reinstatement from room reliability. (A.) Relationship between POCN object reinstatement and room reliability. Regions where room reliability predicted POCN object reinstatement across both guided and free recalls. Objects placed in rooms with the most pre-learning neural stability in these regions were reinstated the most strongly during retrieval. **(B.)** Model comparison results. Regions where room reliability predicted POCN object reinstatement across both guided and free recalls and there was a predictive benefit from participant-specific room reliability. In these regions, the rooms that were most reliable for a specific participant (rather than rooms that were generally reliable across the group) were predictive of object recall for that specific participant. The surface maps presented in **B** show the intersection of the participant-specific models shown in **A** and the regions where there was a significant positive difference in the coefficient of determination between the original participant-specific model and the N-1 group model. Statistical significance for the differences between the coefficients of determination was determined by comparing the differences to a null distribution and FDR-correcting for $q < 0.05$. **(C.)** Controlling for room reinstatement. Left column: Schematic illustrating how room reinstatement evidence in RRCN (during timepoints in which participants verbally recalled a room or its paired-object) was regressed out of room reliability and POCN object reinstatement scores. Room reliability residuals were then correlated at each searchlight with POCN object reinstatement residuals. Right column: Regions where room reliability predicted POCN object reinstatement after controlling for room reinstatement during room and object recall. The surface maps presented in **A** and **C** were statistically thresholded by comparing correlations to a null distribution and then FDR-correcting for $q < 0.05$. All three surface maps are colored based on the magnitude of the z-scored correlation values of the participant-specific model, with blue showing negative and red showing positive relationships, respectively.

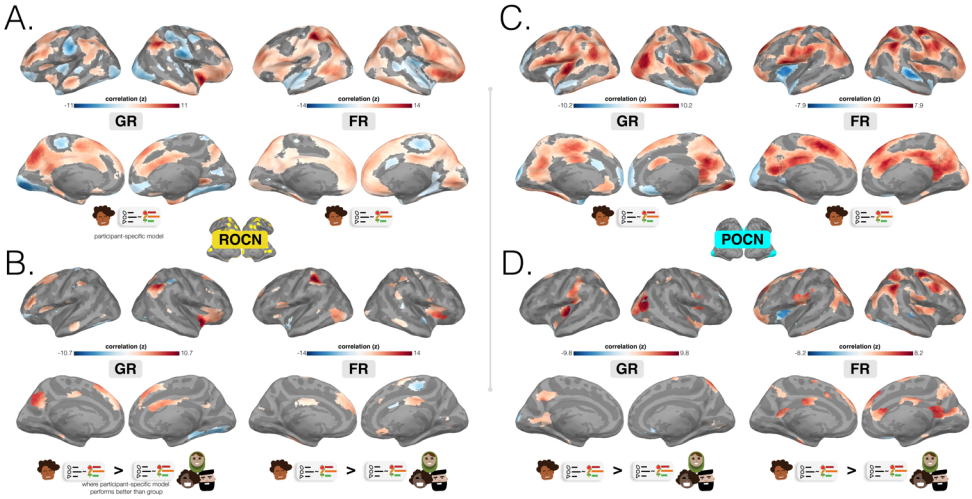

Figure 6–Figure supplement 2. Predicting ROCN and POCN object reinstatement from room reliability. Relationships between room reliability and classifier network object reinstatement evidence. ROCN and POCN results are shown to the left and right of the dividing gray line, respectively. **(A.)** Regions where room reliability predicted ROCN object reinstatement in guided recalls (GR; left) and free recalls (FR; right). **(B.)** Model comparison results. Regions where room reliability predicted ROCN object reinstatement and there was a predictive benefit from participant-specific room reliability, shown separately for guided and free recalls. **(C.)** Regions where room reliability predicted POCN object reinstatement in guided recalls (GR; left) and free recalls (FR; right). **(D.)** Model comparison results. Regions where room reliability predicted POCN object reinstatement and there was a predictive benefit from participant-specific room reliability, shown separately for guided and free recalls. The surface maps presented in A and C were statistically thresholded by comparing correlations to a null distribution and then FDR-correcting for $q < 0.05$. The surface maps presented in **B** and **D** show the intersection of the participant-specific models (**A** and **C**) and the regions where there was a significant positive difference in the coefficient of determination between the original participant-specific model and the N-1 group model. Statistical significance for the differences between the coefficients of determination was determined by comparing the differences to a null distribution and FDR-correcting for $q < 0.05$. All surface maps are colored based on the magnitude of the z-scored correlation values of the participant-specific model, with blue showing negative and red showing positive relationships, respectively.

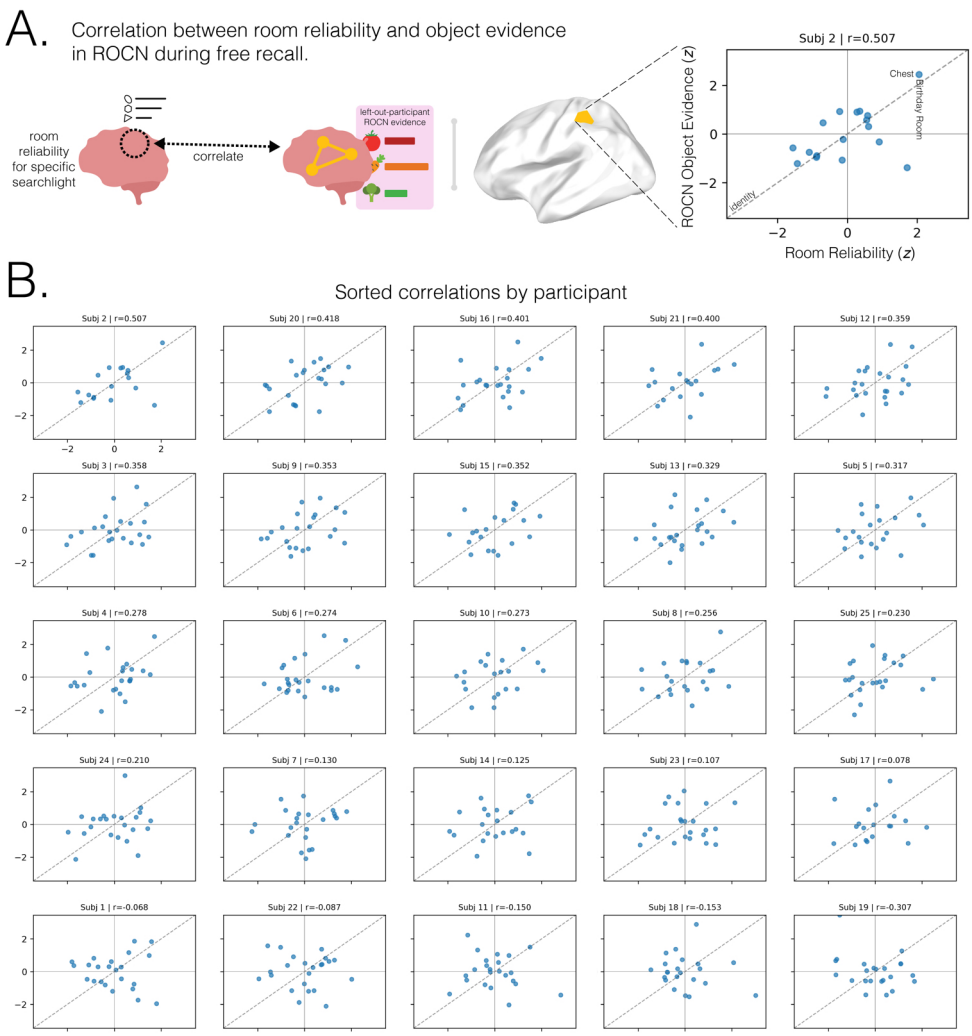

Figure 6–Figure supplement 3. Correlation plots between room reliability and object reinstatement in ROCN for an example searchlight. (A.) Scatterplot illustrating the correlation for an example searchlight and participant. For each participant's free recall (FR) data, we computed the correlation between room reliability and ROCN object reinstatement evidence in the ROCN mask. **(B.)** Individual scatterplots for all 25 participants showing the relationship between room reliability scores and ROCN reinstatement evidence, extracted from the example searchlight seen in **A**. Plots are sorted by participants, moving from left to right in order of highest to lowest correlations, respectively. Dashed lines in scatterplots represent the identity line ($y = x$), and Pearson's r correlation values are reported on the title-line of each plot. Room reliability and ROCN object reinstatement scores were z-scored to facilitate visualization of relationships.