Journal of Experimental Psychology:
Learning, Memory, and Cognition

ISSN: 0278-7393

2022, Vol. 48, No. 7, 929–941
https://doi.org/10.1037/xlm0000894

Spatial Gist Extraction During Human Memory Consolidation

James W. Antony[1], Caroline A. Stiver[2], Kathryn N. Graves[3], Jarryd Osborne[4], Nicholas B. Turk-Browne[3], and Kelly A. Bennion[2]

[1] Princeton Neuroscience Institute, Princeton University
[2] Department of Psychology and Child Development, California Polytechnic State University
[3] Department of Psychology, Yale University
[4] Department of Neuroscience, University of Pittsburgh

Theories of memory consolidation suggest that initially rich, vivid memories become more gist-like over time. However, it is unclear whether gist-like representations reflect a loss of detail through degradation or the blending of experiences into statistical averages, and whether the strength of these representations increases, decreases, or remains stable over time. We report three behavioral experiments that address these questions by examining distributional learning during spatial navigation. In Experiment 1, human subjects navigated a virtual maze to find hidden objects with locations varying according to spatial distributions. After 15 minutes, 1 day, 7 days, or 28 days, we tested their navigation performance and explicit memory. In Experiment 2, we created spatial distributions with no object at their mean locations, thereby disentangling learned object exemplars from statistical averages. In Experiment 3, we created only a single, bimodal distribution to avoid possible confusion between distributions and administered tests after 15 minutes or 28 days. Across all experiments, and for both navigation and explicit tests, representations of the spatial distributions were present soon after exposure, but then receded over time. These findings suggest gist-like representations do not improve over time, helping to clarify the temporal dynamics of consolidation in human learning and memory.

Keywords: episodic memory, false memory, memory consolidation, statistical learning, virtual reality

Supplemental materials: https://doi.org/10.1037/xlm0000894.supp

A central question of memory research is how to resolve the push-and-pull between storing unique yet related episodes (episodic memory) and extracting commonalities across them (gist extraction/statistical learning). For example, whereas in some cases it is important to remember episodes (e.g., where you parked your bike at work today), in other cases it can be helpful to rely on generalized experiences (e.g., where you usually park your bike).

An influential computational model of learning proposed that episodic memory and statistical learning can occur simultaneously because of complementary learning systems in the brain (McClelland et al., 1995). Episodic memories can be formed quickly by medial temporal lobe structures like the hippocampus, whereas statistical learning occurs slowly via accumulated experiences in the neocortex. Repeated interplay between these structures allows for this slow extraction to occur (McClelland et al., 1995). The complementary learning systems framework accords well with theories of memory consolidation, which explain how initially hippocampal-dependent episodic memory traces can be transformed (in at least some form) over time and/or with repeated experience to the neocortex (Nadel & Moscovitch, 1997; Squire et al., 2015; Winocur et al., 2010).

This transformation process aligns with how memories lose detail and specificity over time (Rosenbaum et al., 2000; Wiltgen & Silva, 2007; Winocur et al., 2007). However, according to these models, aspects common across multiple experiences should be better maintained over time because they will be repeatedly reactivated by these later experiences and/or embedded within existing knowledge structures. This idea has garnered support from multiple research domains. The central aspects of stories are less likely to be forgotten than peripheral information (Bartlett, 1932). Lure words that are falsely remembered as a result of being semantically related to studied words get

This article was published Online First July 29, 2021.
James W. Antony https://orcid.org/0000-0003-0656-2170
Caroline A. Stiver https://orcid.org/0000-0003-4563-2282
Kathryn N. Graves https://orcid.org/0000-0001-6701-7681
Kelly A. Bennion https://orcid.org/0000-0002-3301-7744
Our predictions, code, and data can be found here: https://osf.io/mwfq4/wiki/home/.

This work was supported by a CV Starr fellowship to James W. Antony. The authors thank the following individuals (all students of Kelly A. Bennion at California Polytechnic State University, San Luis Obispo) for their assistance with data collection: Fei Ye, Kendra Cornell-Rath, Will Hirsch, Jenny Lee, Cathy Vu, Lauren Bozzetto, Melina Sneesby, Mytien Le, Nancy Lagunas, Aqua Uwadiale, and Hannah Levitt. The authors also thank Daniel Woolley for assistance with the water maze task. The authors declare no competing interests.
Correspondence concerning this article should be addressed to James W. Antony, who is now at Center for Neuroscience, University of California, Davis, 1544 Newton Court, Davis, CA 95618, United States. Email: jantony@princeton.edu

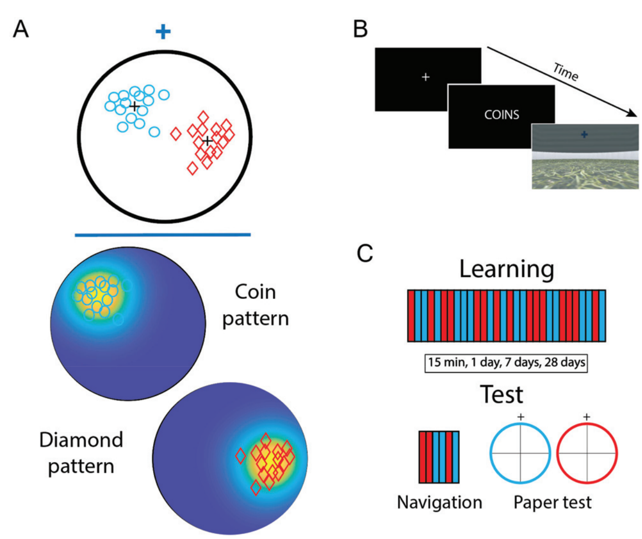
forgotten more slowly over time than words that were actually studied (Thapar & McDermott, 2001). Additionally, category prototypes learned from viewing similar exemplar stimuli are forgotten more slowly over time than exemplars (Posner & Keele, 1970; Strange et al., 1970).

Although gist extraction has been observed across a variety of paradigms and stimuli, the time course over which such representations emerge is less clear. In category learning and false memory paradigms, for example, gist extraction can be detected within minutes of learning (Gallo, 2010; Podell, 1958; Posner & Keele, 1970; Read, 1996; Roediger & McDermott, 1995), suggesting that it occurs either during learning or very quickly thereafter. Relatedly, statistical learning—often thought of as a slow process in theories of memory—can occur very rapidly in minutes (Fiser & Aslin, 2001; Saffran et al., 1996; Turk-Browne et al., 2005), even over multiple interleaved sets of regularities (Gekas et al., 2013). This is consistent with evidence that statistical learning may depend upon the hippocampus (Schapiro et al., 2014; Schapiro et al., 2012; Schapiro et al., 2017), the fast-learning brain system otherwise linked to episodic memory. Relatedly, these literatures support the key ideas of fuzzy trace theory, which suggests that both detailed, episodic and less detailed, gist-like traces are created in parallel during learning, and gist-like traces fade more slowly in memory (Reyna & Brainerd, 1995).

Despite findings in the studies above that these representations emerge rapidly before fading, a recent rodent study showed that gist-like representations improve with longer delays (Richards et al., 2014). Over a series of days, rodents searched for hidden platforms in a circular maze with locations that varied according to an underlying spatial distribution. After 28 days but not 1 day, patterns of rodent search paths were better predicted by the distribution than by individual platforms. These results suggest that as memory for specific details decays with time, gist may be increasingly extracted in the absence of continued experience.

We asked how quickly gist extraction occurs and whether this knowledge then increases, decreases, or remains steady over time. In the first two experiments, subjects (Experiment 1: N = 136; Experiment 2: N = 175) searched a virtual arena for two types of hidden objects: coins and diamonds (see Figure 1A). In Experiment 1, the locations of each type of object varied according to a spatial distribution over the arena. In Experiment 2, each of the two distributions was circular with no object at its mean location, allowing memory for the learned locations to be distinguished from a more gist-like representation of the mean. After this initial learning phase in both experiments, subjects returned for a test session either 15 minutes, 1 day, 7 days, or 28 days later. In this session, we compared trajectories of virtual navigation to the learned locations and distributions to assess spatial memory. After these trials, we also administered a paper test, in which subjects explicitly marked each object location they had encountered. In a third experiment (N = 193), we used only a single distribution (coins) and tested subjects either 15 minutes or 28 days later. Contrary to

Figure 1
Experiment 1 Design



Note. (A, top) Bird's-eye view schematic of the circular environment with object distributions. Subjects were encouraged to orient to the environment using the presence of a blue plus sign on the north wall and a blue south wall. Cyan circles denote coin locations, whereas red diamonds denote diamond locations. Black plus signs denote the means of each distribution. (A, bottom) For each object type, 2D heat maps were created to construct a distributional pattern against which navigation trajectories and paper test responses were measured using KL divergence. (B) Time course of a single trial. Subjects were shown a fixation cross followed by the type of object for which they should search before being placed along the wall of the environment. The lower right-most panel shows a screenshot from the beginning of a sample trial. (C) Overall study procedure. During learning, subjects searched for each object 15 times (30 total) in a random order (in this example, coins are in cyan, diamonds in red). After retention intervals of 15 min, 1 day, 7 days, or 28 days, subjects returned to take a navigation test (three 1-min trials for each object) and an explicit paper test. See the online article for the color version of this figure.

our initial predictions based on rodent findings in Richards et al. (2014), in all three experiments, gist-based representations were learned rapidly and decreased over time.

Experiment 1

We predicted that there would be a dissociation between object exemplar memory (as assessed by the percent of the trajectory spent in the learned locations) and gist-like representations (as assessed by Kullback-Leibler [KL] divergence between a distribution and the search trajectory). In line with the well-documented loss of detail in memory over time, we predicted that memory for specific exemplars would decrease monotonically over time. Inspired by findings from Richards et al. (2014), we predicted that gist extraction would be enhanced by consolidation (i.e., the intervals longer than 15 mins) but also that these representations would fade over time (i.e., by 28 days later; Posner & Keele, 1968; Thapar & McDermott, 2001). Based on these somewhat conflicting findings, we predicted greater distribution-based navigation at the intermediate retention interval conditions (namely, the one-day and seven-day conditions), relative to the 15-min and 28-day conditions, and greater distribution-based navigation for the 28-day than the 15-min condition. We had the same predictions for the paper test.

Method

Subjects

All subjects ($N = 143$; 56 male, 87 female) were undergraduate students with normal or corrected-to-normal vision. Of these subjects, seven (five male, two female) were excluded for not finishing the initial learning phase within the allotted 30-min period; those excluded did not differ significantly in age, $F(1, 141) = .077$, $p = .782$, or gender, $\chi^2(1, \; N = 143) = 3.22$, $p = .073$, from those included in the final analysis. Each subject was randomly assigned to the 15-min, 1-day, 7-day, or 28-day condition, which represented the delay between sessions. At the time subjects signed up for the first session, they did not know when their second session would take place; this information was conveyed to them afterward. If subjects had a scheduling conflict, they were given the opportunity to return for their second session at a different time than previously assigned. We aimed for 30 subjects/condition and stopped new signups after that number had been reached. All procedures were in accordance with the California Polytechnic State University, San Luis Obispo Institutional Review Board. Subjects provided informed consent and earned research credits for their introductory psychology course in exchange for participation.

Stimuli

We used a three-dimensional, circular virtual environment similar to the Morris water maze (Morris, 1984) that was shared from another lab group and used in previous publications (Graves et al., 2020; Woolley et al., 2013; Woolley et al., 2010). The environment was constructed in Blender (www.blender.org) and rendered in MATLAB (ver. 2017a; MathWorks). The environment was altered by adding a grassy texture to the floor. To allow subjects to orient themselves within the environment, the north wall was gray with a blue plus sign, the south wall was blue, and the east and west walls were gray. The radius of the virtual environment was 7.85 arbitrary units (AU), and each object had a radius of 1.5 AU. Following Richards et al. (2014), the locations were determined by sampling the angle from a von Mises distribution ($\pi = 135°$, $\kappa = 3$) and the radius from center from a normal distribution ($M = 3.4$ AU, $SD = 1.13$ AU). To control the complexity of learning the two distributions, the coin and diamond distributions were identical except rotated clockwise 140°. We chose 140° rather than a more regular interval (e.g., 90 or 180) to reduce the likelihood that subjects would be able to learn one distribution with respect to the other using a simple heuristic (e.g., "diamonds = coins + 180°") rather than via experience.

Design and Procedure

Learning Phase. Subjects received instructions that they would be searching within the virtual world (shown from an allocentric viewpoint) for two different types of objects (coins and diamonds). They were told to use the 'I' key to move forward and the 'J' and 'K' keys to move left and right, respectively. They could only move forward, while rotations altered the direction of these forward movements. Subjects were then given a 30-s opportunity to navigate around the virtual world to ensure that they understood how to move around. During the learning phase, subjects engaged in 30 trials (15 coins and 15 diamonds in a randomized order; see Figure 1C). On each trial, they saw a 1.5-s fixation cross, a 2.0-s prompt indicating whether they should look for coins or diamonds, and then, starting from a random place along the edge of the environment, they were given an unlimited amount of time to find the object (see Figure 1B). Upon finding the object, the walls of the environment turned green for 3 s and the program advanced to the next trial. No actual objects were shown within the environment; subjects navigated to locations simply by following the word prompts. This continued until subjects either found all 30 objects or 30 minutes had elapsed. See Figure S1 in the online supplemental materials for learning-related measures.

Test Phase. Subjects returned to the laboratory after their assigned delay for a test in which they were told to navigate to the locations of the objects that they previously found in the learning phase. Retention tests occurred 15 min, 1 day, 7 days, or 28 days later; we included 15 minutes as the earliest delay to prevent the possibility that information could remain in working memory. There were three separate test trials for each object type (in a randomized order), each of which lasted 1 minute. Unbeknownst to the subjects, there were no objects hidden. After the navigation test, they were asked to indicate on separate circular grids from an allocentric viewpoint (Figure 1C) their explicit memory of the coin and diamond locations. Subjects had seen such an allocentric viewpoint once before during the initial instructions. Subjects were asked to indicate the same number of locations as they had learned for each object type (15) and were encouraged to guess if they could not remember. Finally, they completed a posttest questionnaire with questions about their strategy for finding the objects.

Spatial Learning Measurements

We had two main variables of interest during navigation test trials. First, to assess memory for individually learned locations, we calculated the percentage of the trajectory spent within object locations. This analysis separated each moment within the trajectory as either inside or outside any learned location and did not double count if the subject was in a location that overlapped with two objects. Because this metric does not indicate how memory performance differs from random navigation (no memory), we also calculated these percentages after rotating trajectories every 10° within the full 360° space (creating scrambled distributions). We then subtracted the mean percentage of the rotated (scrambled) distributions from the true percentage to assess whether performance differed from chance. We used this rotation method rather than simply dividing the locations by the full arena to account for the possibility that subject-specific navigational strategies could be biased toward locations that resemble the learned locations (e.g., at a particular radius from the center), which could result in above-chance results without necessarily indicating any true memory for the experience. We do note, however, that with some increments (e.g., 140°), these distributions occasionally will be rotated onto each other, but overall, the effect of such rotations will be averaged out by using the full 360° space. Any such alignment on a subset of increments would in fact diminish the likelihood of finding a difference between congruent and incongruent conditions. Finally, we calculated separate measures for congruent memory (e.g., how much time subjects spent in coin locations when given a coin prompt) and incongruent memory (e.g., how much time subjects spent in coin locations when given a diamond prompt).

Second, to assess gist-based memory, we asked how well the subject's trajectory matched each spatial distribution. To do this, we took "snapshots" of the subject's location five times per second and used a bivariate kernel density estimator to create a two-dimensional spatial pattern. We similarly created a pattern with the true object locations. We chose a relatively wide kernel to capture broad regularities within each pattern; visualization of patterns using different smoothing kernels are shown in Figure S2 in the online supplemental materials. We then measured the pattern match by calculating KL divergence, as used by Richards et al., 2014, which sums the differences at each pixel within the two-dimensional space. Note that lower KL divergence indicates a better pattern match. After creating the pattern, we added .001 to each pixel of the distribution before summing the entire distribution to 1 to prevent issues related to individual pixels within the distribution containing a 0 denominator in the KL divergence formula. We calculated these values for the coin and diamond distributions separately. Because this metric does not account for random navigation, we also calculated KL divergence after rotating trajectories every 10° to assess chance performance.

After the navigation tests, we measured memory via a paper test (see Figure 1C). The purpose of the paper test was to assess whether our effects apply to a separate explicit test where subjects can map out their memories in a less time-constrained fashion. A single rater scored each set of x and y coordinates by hand. The paper circles had radii of 6.5 cm and coordinates were scored to 1 mm accuracy using a ruler, after which they were converted to arbitrary units representing the virtual environment. For this test, we assessed individual memory by matching guessed locations to true object locations. Since there was no way to verify which guess was intended for which location, we matched guessed and true locations over a series of steps. First, we created a full matrix of all the distances between guessed and true locations. Then, we chose the minimum distance from the matrix and dropped that guess and true location out of the analysis. Then, we created a new full matrix and repeated this process until we were left with one guessed and one true location for that distribution. Individual memory from this test was operationalized as the average distance from these guesses, as perfect memory would constitute placing a guess at each learned location, and thus zero spatial error. We assessed gist-based memory by creating a 2-dimensional spatial pattern out of the guessed locations and measuring KL divergence from the true spatial pattern, similar to the process outlined above during navigation.

Statistical Analyses

To verify that learning occurred, we used a paired samples t test to compare the amount of time it took to find objects on the average of the first versus last two trials of each type. We also aimed to quantify whether subjects used more of an exemplar-based versus gist-based strategy during learning. To do this, we used the last learned location as the exemplar from which subjects would most likely rely. We compared the percentage of time spent in the last learned location and the experienced mean (the mean of all trials of that type up to that point in learning) of that distribution via paired samples t-tests.

For the main memory analyses, we submitted the different metrics to one-way ANOVAs, with retention interval (15 min, 1 day, 7 days, 28 days) as a between-subjects factor. Principally, we considered congruent minus incongruent metrics to assess memory, using percentage of time within object locations for individual memory and KL divergence for gist-based memory. For any significant results, we conducted follow-up independent samples t-tests between each comparison.

Data Availability

Analysis code and data sets from this study were made available to the reviewers during the review process and will be made publicly available at the time of publication on the Open Science Framework website (https://osf.io/mwfq4/wiki/home/).

Results and Discussion

Learning

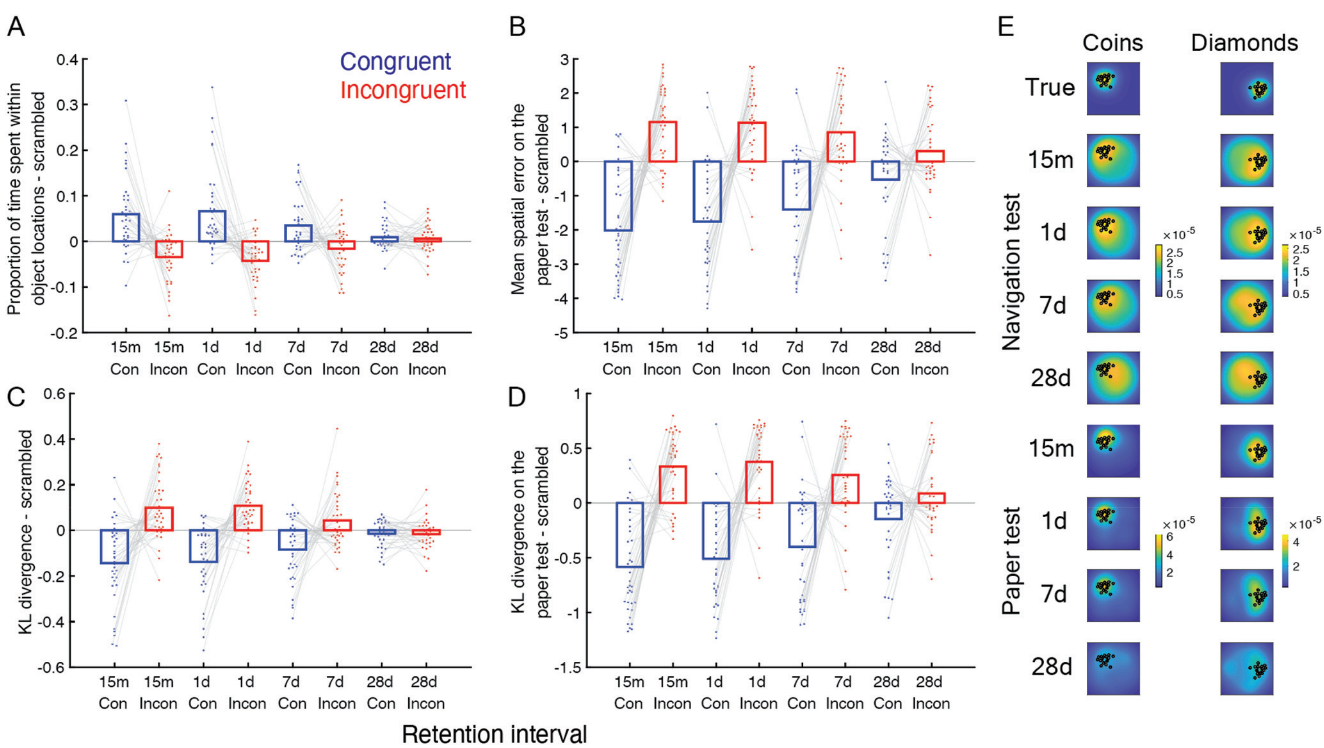
Subjects showed consistent learning in the initial phase across all delay conditions, as would be expected (because the retention interval started after this phase). Learning was evidenced by reduced time to find target objects across the 15 trials (see Figure S1A in the online supplemental materials; first vs. last two trials: coins first = 41.3 ± 3.0; last = 22.6 ± 1.7; $t[135] = 5.5$, $dz = .47$, $p < .001$; diamonds first = 47.2 ± 3.7; last = 23.2 ± 1.5; $t[135] = 6.3$, $dz = .54$, $p < .001$). We also assessed their strategy during learning by calculating the amount of time they spent in the last learned location for that object type and in the experienced mean of the corresponding distribution. Overall, subjects spent more of their time in the previously learned location than the experienced mean location (see Figure S1B in the online supplemental materials; previous: 5.3 ± .2%; M: 4.7 ± .2%; $t[135] = 5.3$, $dz = .46$, $p < .001$).

Retention

We predicted that performance for individual exemplars would decrease over time. In line with this prediction, we found that the percentage of time spent in learned object locations decreased over time on the navigation test (for congruent—incongruent: 15m: 9.4 ±2.2%; 1d: 10.9 ± 2.2%; 7d: 5.1 ± 1.8%; 28d: .3 ± .9%; F[3, 133] = 14.59, p < .001; see Figure 2A). Similarly, the average spatial error on the explicit paper test increased over time (for congruent—incongruent: 15m: −3.17 ± .42 AU; 1d: −2.89 ± .46 AU; 7d: −2.26 ± .51 AU; 28d: −.83 ± .43 AU; F[3, 133] = 12.93, p < .001; see Figure 2B). Follow-up analyses on the navigation test indicated differences between 15 min versus 28 days, t(66) = 3.72, d = .92, p < .001, 1 day versus 7 days, t(66) = 2.01, d = .49, p = .048, 1 day versus 28 days, t(64) = 4.32, d = 1.08, p < .001, and 7 days versus 28 days, t(64) = 2.28, d = .57, p = .026. On the paper test, follow-up analyses indicated differences between 15 min versus 28 days, t(66) = 3.76, d = .91, p < .001, 1 day versus 28 days, t(64) = 3.14, d = .77, p = .003, and 7 days versus 28 days, t(64) = 2.05, d = .51, p = .044.

We predicted more gist-based navigation at the 1-day and 7-day intervals than at the 15-min or 28-day intervals. We calculated this by measuring KL divergence of navigation trajectories for each object type (coins or diamonds) compared with each distribution (coins or diamonds), and then grouped the congruent (e.g., coin search for coin distribution) versus incongruent (e.g., coin search for diamond distribution) combinations. We then contrasted congruent minus incongruent KL divergence differences for each retention interval, using a one-way ANOVA with 4 levels (15 min, 1 day, 7 days, 28 days). Contrary to these predictions, memory for the distributions decreased over time on the navigation test (for congruent—incongruent: 15m: −.24 ± .05 AU; 1d: −.25 ± .04 AU; 7d: −.13 ± .04 AU; 28d: .002 ± .02 AU; F[3, 133] = 20.6, p < .001; see Figure 2C) and the explicit paper test (for congruent—incongruent: 15m: −.94 ± .13 AU; 1d: −.90 ± .13 AU; 7d: −.67 ± .15 AU; 28d: .24 ± .13 AU; F[3, 133] = 13.8, p < .001; see Figure 2D). Follow-up analyses of the navigation test indicated differences between 15 min versus 28 days, t(66) = 4.30, d = 1.07, p < .001, 1 day versus 28 days, t(64) = 5.34, d = 1.33, p < .001, and 7 days versus 28 days, t(64) = 2.66, d = .66, p = .010, and marginally significant differences between 15 min versus 7 days, t(68) = 1.70, d = .41, p = .094, and 1 day versus 7 days, t(66) = 1.96, d = .47, p = .055. On the paper test, follow-up analyses indicated differences between 15 min versus 28 days, t(66) = 3.82, d = .93, p < .001, 1 day versus 28 days, t(64) = 3.5, d = .86, p < .001, and 7 days versus 28 days, t(64) = 2.09, d = .52, p = .041. To verify that this effect was not driven by the incongruent combinations, we repeated these analyses considering the congruent combinations alone. They again showed a significant decrease over time on both the navigation (15m: −.14 ± .03 AU; 1d: −.14 ± .03 AU; 7d: −.08 ± .02 AU; 28d: −.02 ± .01 AU; F[3, 133] = 15.82, p < .001) and paper tests (15m: −.58 ± .07 AU; 1d: −.51 ± .08 AU; 7d: −.40 ± .09 AU; 28d: −.15 ± .07 AU;

Figure 2
Experiment 1 Results



Note. (A and B) Learned location memory was plotted as the percentage of time within object locations for the navigation test (A) and mean spatial error for the paper test (B) for each retention interval and congruency type (e.g., congruent indicates a coin search prompt against a coin distribution). (C and D) Gist-based memory was plotted as the KL divergence for the navigation test (C) and KL divergence for the paper test (D) for each retention interval and congruency type. (E) True and average distribution patterns for all retention intervals of attended locations on the navigation test (top) and guessed locations on the paper test (bottom) for coin (left) and diamond (right) distributions. Learned locations are represented as small circles for visibility, with the coin and diamond mean locations as larger circles. See the online article for the color version of this figure.

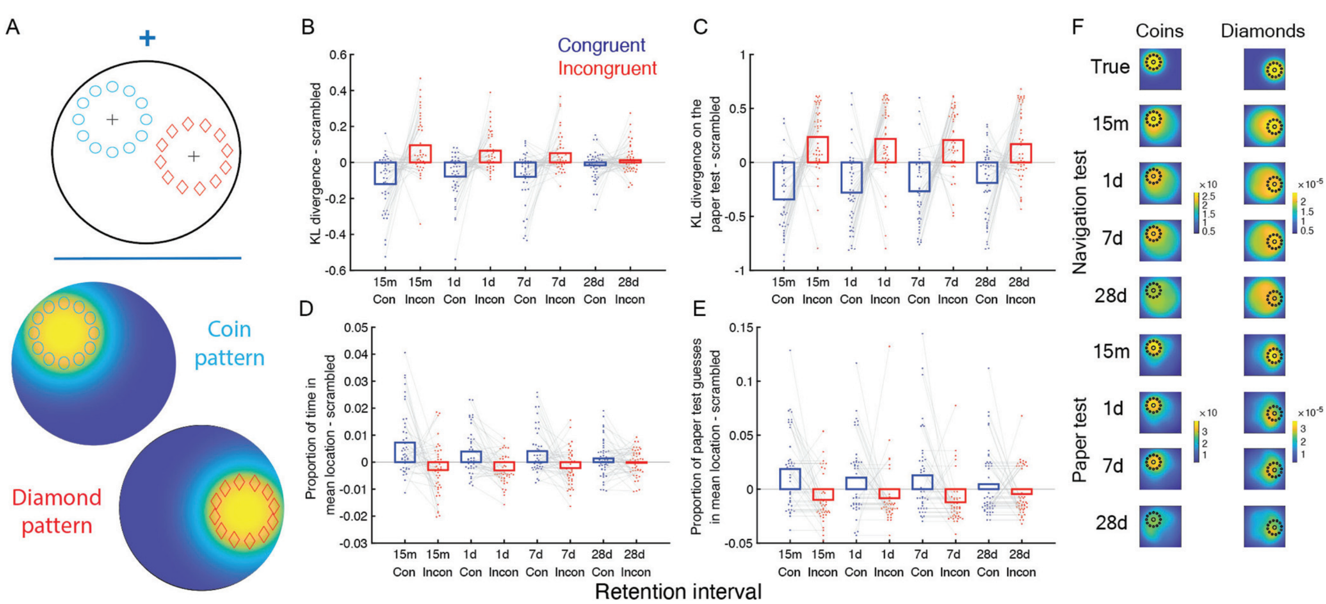
F[3, 133] = 16.43, $p < .001$). Contrary to our predictions, gist-like representations decreased over time, with no improvement at any delayed retention interval.

Another way to conceptualize gist in this paradigm is that learned object locations may be maintained but their type (i.e., coins vs. diamonds) forgotten or swapped. In this context, gist would correspond to the spatial distribution pooling across all locations blind to type (i.e., "unlabeled"). We therefore calculated KL divergence adding both types of searches against both distributions (i.e., coin search for coin distribution + coin search for diamond distribution, diamond search for coin distribution + diamond search for diamond distribution) and measured how they changed over time and differed from chance performance (i.e., with scrambled distributions rotated in 10-degree increments). Here we found a dissociation between the navigation and paper tests: there was an effect of time on the paper test (15m: −.25 ± .04 AU; 1d: −.13 ± .03 AU; 7d: −.14 ± .03 AU; 28d: −.06 ± .03 AU; F[3, 133] = 13.9, $p < .001$) but no longer on the navigation test (15m: −.04 ± .02 AU; 1d: −.03 ± .02 AU; 7d: −.04 ± .01 AU; 28d: −.03 ± .01 AU; F[3, 133] = .16, p = .70). On the paper test, follow-up analyses showed better memory at earlier intervals between 15 min versus 1 day, t(68) = 2.34, d = .56, p = .02, 15 min versus 7 days, t(68) = 2.10, d = .50, p = .04, 15 min versus 28 days, t(66) = 4.02, d = .99, $p < .001$, and marginally between 7 days versus 28 days, t(64) = 1.86, d = .46, p = .067. In assessing whether these pooled KL divergence measures differed from chance, all retention intervals were significant on the navigation test except the one-day delay, 15m: t(35) = 2.86, dz = .48, p = .007; 1d: t(33) = 1.59, dz = .27, p = .12; 7d: t(33) = 3.70, dz = .63, $p < .001$; 28d: t(35) = 2.39, dz = .42, p = .02, and all retention intervals were significant on the paper test, 15m: t(35) = 6.82, dz = 1.14, $p < .001$; 1d: t(33) = 3.98, dz = .68, $p < .001$; 7d: t(33) = 4.14, dz = .71, $p < .001$; 28d: t(35) = 2.05, dz = .36, p = .048. Together, these results provide evidence that unlabeled gist-like representations are either stable or worsen over time; however, there is no evidence that these representations improve as we hypothesized or as was found by analogy in Richards et al. (2014).

Experiment 2

In Experiment 1, some individual object locations overlapped with the mean location of the distribution, making it difficult to disentangle memory for individual exemplars versus gist-like statistical averages. We addressed this issue in Experiment 2 by creating new distributions of coins and diamonds such that their locations were arranged in a circle around a mean location at which no object could be found (see Figure 3A). In this way, any memory for objects at the mean location would provide clear evidence for a gist-like representation (see Smith & Minda, 2002, for a conceptually similar paradigm). Here we again predicted that these gist representations of the mean location would be stronger for one-day and seven-day retention intervals than for 15-min and 28-day retention intervals. Nevertheless, an alternative possibility based on Experiment 1 is that gist-like representations would decrease in strength over time.

Figure 3
Experiment 2 Object Locations and Results



Note. (A, top) A bird's-eye view schematic of the circular environment in Experiment 2 shows object distributions, with 12 cyan circles representing coins and 12 red diamonds representing diamonds. (A, bottom) Patterns used to assess KL divergence for each object type are shown. (B and C) Gist-based memory was plotted for the navigation test (B) and paper test (C) for each retention interval and congruency type. (D and E) Mean (false) location memory as assessed by percentage of time within mean location is plotted for the navigation test (D) and percent of responses within the mean location for the paper test (E) for each retention interval and congruency type. (F) True and average distribution patterns for all retention intervals of attended locations on the navigation test (top) and guessed locations on the paper test (bottom) for coin (left) and diamond (right) distributions. Learned locations are represented as small circles for visibility, with the coin and diamond mean locations as larger circles. See the online article for the color version of this figure.

Method

Subjects

All subjects (N = 279; 86 male, 193 female) were undergraduate students with normal or corrected-to-normal vision. Of these subjects, 104 (23 male, 81 female) were excluded here for not finishing the learning phase within the allotted 30-min period. The excluded participants did not significantly differ in age from those included in final analyses, $F(1, 277) = .58$, $p = .446$, although females were more likely than males to not finish the learning phase and therefore be excluded, $\chi^2(1, N = 279) = 5.90$, $p = .015$. The high exclusion rate likely stemmed from smaller object sizes (see Stimuli), making the task more difficult. Similar to Experiment 1, we aimed for 30 subjects/condition and stopped new sign-ups after that number had been reached.

Stimuli

The stimuli were mostly the same as Experiment 1, with three modifications: object sizes were reduced from a radius of 1.5 AU to 1 AU to eliminate overlap with the mean location; the distributions (140° apart) were created by rotating locations around the center of a circle (2.7 AU radius) in 12 equal increments (see Figure 3A), and there were fewer objects overall (12 coins and 12 diamonds in a randomized order). These locations ensured that the center of the circle was the exact mean of the learned locations and aligned with the peak of the heat map for the KL divergence analyses.

Design and Procedure

Learning Phase. All procedures were the same as in Experiment 1, with the exception that there were fewer trials (24 total) to account for anticipated increased learning time owing to the smaller object sizes (reduced radius from 1.5 to 1 AU).

Test Phase. All procedures were the same as in Experiment 1.

Statistical Analyses

To compute patterns, we used the same smoothing kernel to that in Experiment 1; visualization of patterns using different smoothing kernels are shown in Figure S3 in the online supplemental materials. The majority of analyses were identical to Experiment 1. To calculate mean memory in Experiment 2, we submitted percentage of time spent in the congruent versus incongruent mean location to a mixed, 2 (congruency: congruent mean, incongruent mean) $\times$ 4 (retention interval: 15 min, 1 day, 7 days, 28 days) ANOVA. This was performed to specifically show that for this metric, there was above-chance memory, representing the statistical average of the distribution.

Simulation

We aimed to confirm that gist-like and individual memory could indeed be differentiated in this study. To accomplish this, we created two sample subjects (S1 and S2) that performed the paper spatial memory task in Experiment 2. S1 made spatial guesses within a very small distance (less than the platform radius) of the statistical mean location. S2 made spatial guesses within identically small distances of the individual platform locations. We calculated both KL divergence and spatial error from the individual locations by subtracting consistent minus inconsistent trials. Better performance on these metrics is reflected by lower (more negative) numbers. S1 (responding within the mean location) had KL divergence from the distribution centered at the mean of -1.61 and spatial error from individual locations of -5.58. This subject had 100% of responses within the mean location. S2 (responding within learned platform locations) had KL divergence of -1.37, individual memory of -7.13, and 0% of responses within the mean location. Therefore, KL divergence was more sensitive to behavior based on the mean than behavior based on individual locations, and the proportion of guesses in the mean location was completely dissociable based on these behaviors.

Results and Discussion

Learning

Subjects demonstrated learning across the 12 trials of the two object types (see Figure S1C in the online supplemental materials; first versus last two trials: coins first = 57.3 $\pm$ 3.3; last = 47.0 $\pm$ 2.7; t[174] = 2.3, dz = .18, p = .02; diamonds first = 62.3 $\pm$ 3.7; last = 47.3 $\pm$ 2.9; t(174) = 3.0, dz = .23, p = .003). Similar to Experiment 1, subjects spent more time in the previously learned than the experienced mean location (see Figure S1D in the online supplemental materials; previous: 2.6 $\pm$ .08%; M: 2.4 $\pm$ .08%; t[174] = 3.6, dz = .21, p = .006).

Retention

We predicted that performance for individual exemplars would decrease over time. In line with this prediction, and similar to Experiment 1, we found that the percentage of time spent in learned object locations decreased over time on the navigation test (for congruent—incongruent: 15m: 7.1 $\pm$ 1.4%; 1d: 4.4 $\pm$ 1.1%; 7d: 4.7 $\pm$ 1.2%; 28d: .5 $\pm$ .7%; F[3, 172] = 13.63, $p < .001$). Similarly, the average spatial error on the explicit paper test marginally increased over time (for congruent—incongruent: 15m: $-2.31 \pm .37$ AU; 1d: $-2.0 \pm .41$ AU; 7d: $-1.91 \pm .36$ AU; 28d: $-1.39 \pm .38$ AU; F[3, 172] = 2.83, p = .09). Follow-up analyses on the navigation test indicated differences between 15 min versus 28 days, $t(87) = 4.14$, $d = .86$, $p < .001$, 1 day versus 28 days, $t(88) = 2.85$, $d = .60$, $p = .005$, and 7 days versus 28 days, $t(88) = 2.88$, $d = .60$, $p = .005$.

We also predicted that with a clearer distinction between the individual object locations and the mean location, subjects might show the hypothesized increase in gist representations at the one-day and seven-day intervals relative to the 15-min and 28-day intervals. Again contrary to these predictions, memory for the distributions (measured as congruent $-$ incongruent KL divergence) decreased over time on the navigation test (15m: $-.22 \pm .04$ AU; 1d: $-.14 \pm .03$ AU; 7d: $-.13 \pm .03$ AU; 28d: $-.03 \pm .02$ AU; F[3, 172] = 13.59, $p < .001$; see Figure 3B) and marginally decreased on the explicit paper test (15m: -0 59 $\pm$ .09 AU; 1d: $-.51 \pm .10$ AU; 7d: $-.48 \pm .09$ AU; 28d: $-.36 \pm .09$ AU; F[3, 172] = 2.74, p = .10; see Figure 3C). Follow-up analyses on the navigation test indicated differences between 15 min versus 28 days, $t(87) = 3.72$, $d = .78$, $p < .001$, 1 day versus 28 days, $t(88) = 2.90$, $d = .61$, $p = .005$, and 7 days versus 28 days, $t(88) = 2.47$, $d = .52$, $p = .016$. We again repeated the analysis for the congruent distribution alone, which showed a significant decrease over time on both the navigation (15m: $-.12 \pm .02$ AU; 1d: $-.08 \pm .02$ AU; 7d: $-.08 \pm .02$ AU; 28d: $-.02 \pm .01$ AU; F[3, 172] = 13.6,

$p < .001$) and paper tests (15m: $-.34 \pm .05$ AU; 1d: $-.28 \pm .05$ AU; 7d: $-.27 \pm .05$ AU; 28d: $-.19 \pm .05$ AU; $F[3, 172] = 4.6$, $p = .03$).

As in Experiment 1, we calculated the amount of time spent in the unlabeled, pooled distribution of locations blind to object type. We again found a dissociation between the navigation and paper tests: we found a significant effect of time on the paper test (15m: $-.11 \pm .02$ AU; 1d: $-.06 \pm .01$ AU; 7d: $-.06 \pm .02$ AU; 28d: $-.02 \pm .02$ AU; $F[3, 172] = 8.9$, $p = .003$) but not the navigation test (15m: $-.03 \pm .01$ AU; 1d: $-.01 \pm .01$ AU; 7d: $-.03 \pm .01$ AU; 28d: $-.004 \pm .007$ AU; $F[3, 172] = 1.4$, $p = .24$). Follow-up analyses of the paper test showed a difference between 15 min versus 28 days, $t(87) = 2.86$, $d = .61$, $p = .005$, and a marginal difference between 15 min versus 1 day, $t(83) = 1.68$, $d = .36$, $p = .097$. In assessing whether these pooled KL divergence measures differed from chance, the 15-min and seven-day intervals were significant on the navigation test, 15m: $t(41) = 2.14$, dz = .33, $p = .03$; 1d: $t(42) = 1.63$, $dz = .25$, $p = .11$; 7d: $t(42) = 2.41$, $dz = .37$, $p = .02$; 28d: $t(46) = .48$, $dz = .07$, $p = .63$, whereas the 15-min, 1-day, and 7-day intervals were significant on the paper test, 15m: $t(41) = 4.64$, $dz = .72$, $p < .001$; 1d: $t(42) = 4.54$, $dz = .69$, $p < .001$; 7d: $t(42) = 2.80$, $dz = .43$, $p = .008$; 28d: $t(46) = .99$, $dz = .14$, $p = .33$.

We were particularly interested in whether subjects would falsely remember finding an object in the statistical average of each distribution (that is, the center of the circle). Therefore, we calculated the percentage of time spent in this location on the navigation test and the percent of guesses made in this location on the paper test. We submitted these measures to mixed 2 (congruency: congruent mean, incongruent mean) $\times$ 4 (retention interval: 15 min, 1 day, 7 days, 28 days) ANOVAs. On the navigation test, we found a significant main effect of congruency (congruent: $.41 \pm .07\%$; incongruent: $-.21 \pm .05\%$; $F(1, 171) = 40.6$, $p < .001$; see Figure 3D). There was no main effect of retention interval (15 m: $2.2 \pm .08\%$; 1 day: $.4 \pm .08\%$; 7 days: $.08 \pm .07\%$; 28 d: $.03 \pm .06\%$; $F(3, 171) = 1.74$, $p = .19$), but there was a congruency by retention interval interaction, $F(1, 171) = 9.61$, $p = .002$. Follow-up analyses (on congruent—incongruent percentages) indicated that false memory for the center location decreased between 15 min versus 28 days, $t(87) = 2.83$, $d = .59$, $p = .006$, 1 day versus 28 days, $t(88) = 2.76$, $d = .58$, $p = .007$, and 7 days versus 28 days, $t(88) = 2.10$, $d = .44$, $p = .04$. On the paper test, we similarly found a main effect of congruency (congruent: $1.2 \pm .3\%$; incongruent: $-.9 \pm .2\%$; $F(1, 171) = 33.9$, $p < .001$; see Figure 3E). There was no main effect of retention interval (15 m: $.4 \pm .3\%$; 1 day: $.2 \pm .3\%$; 7 days: $.08 \pm .3\%$; 28 d: $.2 \pm .3\%$; $F(3, 171) = 1.08$, $p = .30$), and there was a marginally significant congruency by retention interval interaction, $F(1, 171) = 3.02$, $p = .08$. Follow-up analyses (on congruent—incongruent percentages) indicated that false memory for the center location decreased between 15 min versus 28 days, $t(87) = 2.19$, $d = .46$, $p = .03$, and marginally between 7 days versus 28 days, $t(88) = 1.69$, $d = .35$, $p = .095$.

Finally, we contrasted representations of the mean versus the learned locations by calculating the relative proportion of time spent in each on the navigation test or relative proportion of guesses made on the paper test. We performed this analysis by dividing mean location occupancy by the sum of mean test and learned location occupancy, and we calculated it for the consistent locations only (for example, only considering the mean coin location on coin probe trials). We submitted these measures to one-way ANOVAs with 4 levels of retention interval (15 min, 1 day, 7 days, 28 days). On the navigation test, we found a marginally significant effect of retention interval, $F(1, 173) = 3.3$, $p = .07$, with a trend toward decreasing levels of relative mean versus learned location guesses over time (15 m: $10.1 \pm .5\%$; 1 day: $9.5 \pm .6\%$; 7 days: $9.1 \pm .5\%$; 28 d: $8.8 \pm .6\%$). For the paper test, we removed 21 subjects who had no guesses in either the mean or learned locations. We found no significant effect of retention interval, $F(1, 152) = 2.6$, $p = .11$, although this test also showed a quantitative decrease in relative mean guesses over time (15 m: $11.8 \pm 2.2\%$; 1 day: $7.9 \pm 1.5\%$; 7 days: $8.3 \pm 1.7\%$; 28 d: $7.3 \pm 1.5\%$). Overall, these results suggest that gist memory—here represented by common patterns and false memory for unlearned, central information—decreases over time.

Experiment 3

How can we reconcile the results of Experiments 1 and 2 with the compelling finding of Richards et al. (2014) that gist-based memory increases over time? In Experiment 3, we tackle a couple of methodological differences that might explain this discrepancy.

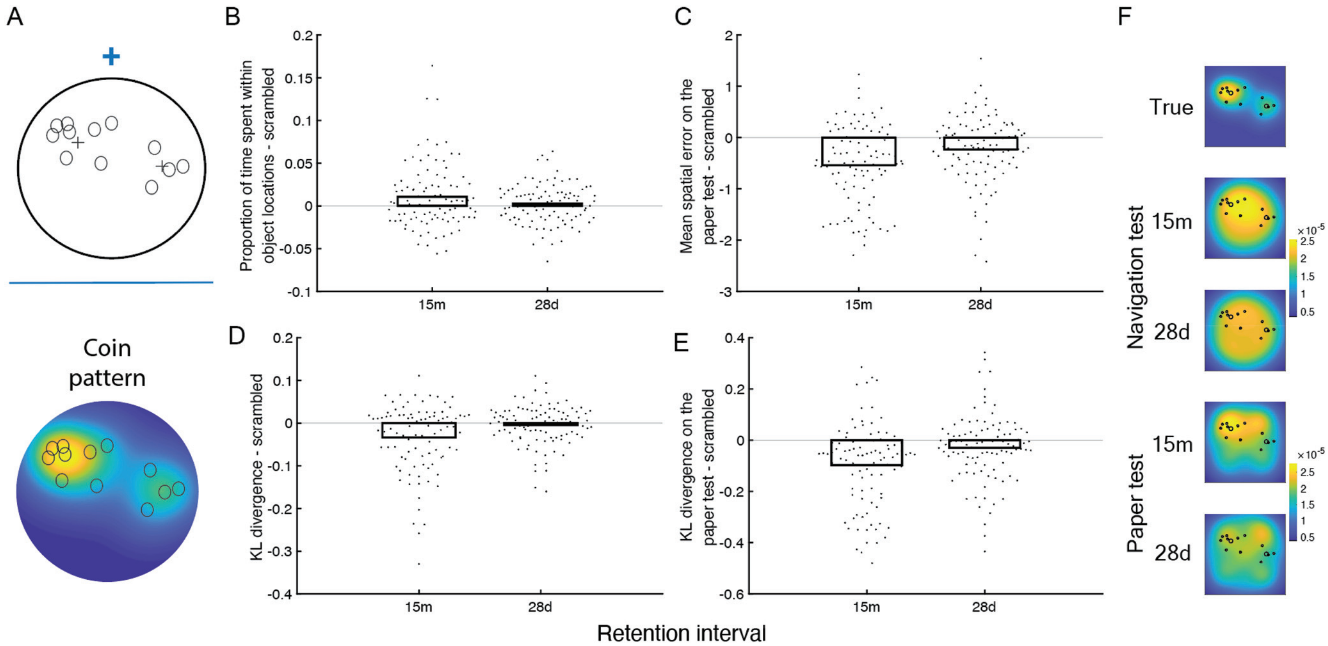
First, in Experiments 1 and 2, subjects learned two spatial distributions, whereas in Richards et al. (2014), only one distribution was learned. We explored this difference above by conducting analyses of whether subjects had an unlabeled gist-like representation for locations pooled across object types. This conceptually resembles having a single distribution. This did impact the results of both experiments by eliminating the decrease of gist representations over time during navigation (though the decrease remained in the paper tests). Note that neither these pooled analyses nor any of the other analyses uncovered an increase in gist representations over time. Nevertheless, there could be an unforeseen consequence to using two distributions rather than one, so here we employed a single distribution.

Second, in Richards et al. (2014) trials were fixed in a particular order, whereas in our study this order was randomized. Therefore, here we manipulated trial order by specifically altering the location of the last learning event.

Third, there are obvious methodological differences between running experiments with rodents and humans. One possible factor affecting our results is that we gave instructions at the final test to navigate toward previously learned locations. Here we explore whether instead asking subjects to navigate to a new location from the same underlying distribution would produce different results. Although rodents could not be instructed in either way, they may have naturally pursued such a strategy.

To summarize the new design, subjects learned a single, bimodal distribution (coin only) with twice as many learning episodes in one mode than the other (resembling the weighted bimodal distribution experiments of Richards et al., 2014; see Figure 4A). We ran six groups of subjects, half of which took a test after 15 minutes and half after 28 days, following findings in Experiments 1 and 2 that there were no U-shaped effects on retention but rather gradual changes over time. We manipulated trial order by having two groups of subjects learn the last location in the more likely mode (15m, major mode: $15m_M$; 28d, major mode: $28d_M$) and another two groups in the less likely mode (15m, minor mode: $15m_m$; 28d, minor mode: $28d_m$; similar to Richards et al., 2014). Additionally, for another two groups (who both had the last location in the less likely mode), we explored whether orienting subjects toward distributions modified performance (15m,

Figure 4
Experiment 3 Object Locations and Results



Note. (A, top) Bird's-eye view schematic of the circular environment shows a bimodal distribution of 12 circles (black). Black plus signs denote the means of each mode. (A, bottom) Patterns used to assess KL divergence are shown. (B and C) Learned location memory for the navigation test (B) and mean spatial error for the paper test (C) are plotted for the 15-min and 28-day intervals. (D and E) Gist-based memory is plotted for the navigation (D) and paper test (E) for the 15-min and 28-day intervals. (F) True and average distribution patterns are shown for both retention intervals of attended locations on the navigation test (top) and guessed locations on the paper test (bottom). Learned locations are represented as small circles for visibility, with the major and minor mode means as larger circles. See the online article for the color version of this figure.

minor mode, distributional instructions: $15m_{mD}$; 28d, minor mode, distributional instructions: $28d_{mD}$). Principally, we were interested in whether these differences could account for the discrepancy between our results and those of Richards et al. (2014).

Method

Subjects

All subjects (N = 204; 79 male, 124 female, 1 nonbinary) were undergraduate students with normal or corrected-to-normal vision. Of these subjects, 11 (one male, 10 female) were excluded here for not finishing the learning phase within the allotted 30-min period; those excluded did not differ in age, $F(1, 202) = .15$, $p = .702$, or gender, $\chi^2(2, N = 204) = 4.43$, $p = .109$) from those included in the final analysis. Similar to prior experiments, we aimed for 30 subjects/condition and stopped new signups after that number had been reached.

Stimuli

The stimuli were the same size as in Experiment 2 (radius of 1 AU). The single bimodal distribution was created from 12 locations in Experiment 1, eight exemplars from the coin distribution (major mode) and four exemplars from the diamond distribution (minor mode; Figure 4A).

Learning Phase

All procedures were the same as in Experiments 1 and 2, except that there were only 12 trials total (all coin trials). Every three trials included two trials from the major mode and one trial from the minor mode, in a random order within each set (except for the final triplet). Within the final triplet, four groups of subjects finished learning in the minor mode ($15m_m$, $28d_m$, $15m_{mD}$, $28d_{mD}$) and two groups finished in the major mode ($15m_M$, $28d_M$).

Test Phase

All procedures were the same as in Experiments 1 and 2, with two exceptions. First, we probed memory in the navigation test with three one-minute search trials and in the paper test with only a single circle. Second, we gave different navigation instructions to two groups of subjects who finished learning in the minor mode ($15m_{mD}$, $28d_{mD}$). They were told to look for locations that may not be identical to previously learned objects but would be drawn from the same distribution.

Memory Measurements

All measurements were the same as in Experiments 1 and 2, except that we could no longer compare congruent and incongruent distributions because there was only one object type. During learning, we compared the percentage of time spent in the last learned location and the experienced mean (the mean of all trials of that type up to that point in learning) of both the major and minor modes of the distribution via paired samples t-tests.

Statistical Analyses

To compute patterns, we used the same smoothing kernel to that in Experiments 1 and 2; visualization of patterns using different smoothing kernels are shown in Figure S4 in the online supplemental materials. The data were analyzed in the same manner as in Experiments 1 and 2. To compare across order and instruction conditions, we submitted the memory measurements to 2 (retention interval: 15 min, 28 days) × 3 (learning condition: last location in major mode, last location in minor mode, last location in minor mode with distributional instructions) ANOVAs.

Results and Discussion

Learning

Subjects demonstrated learning across the 12 trials (see Figure S1E in the online supplemental materials; search time for the first versus last two trials: first = 86.2 ± 7.5 s; last = 60.8 ± 3.3 s; $t(192) = 3.1$, $dz = .22$, $p = .002$). They spent more time in the last learned location (2.2 ± .08%) and the experienced major mode (2.0 ± .1%) than the experienced minor mode mean (1.6 ± .07%; last learned versus minor mode: $t(192) = 6.5$, $dz = .46$, $p < .001$; major mode versus minor mode: $t(192) = 4.2$, $dz = .30$, $p < .001$). There was no significant difference between the last learned and the major mode mean, $t(192) = 1.28$, $dz = .09$, $p = .20$ (see Figure S1F in the online supplemental materials).

Retention

We predicted that performance for individual memories would decrease over time. In line with this prediction, and similar to the previous experiments, the percentage of time spent in learned object locations marginally decreased over time on the navigation test (main effect of retention interval: 15m = 1.1 ± .4%; 28d = .2 ± .2%; $F(1, 190) = 3.20$, $p = .08$; no main effect of learning condition: major mode = .8 ± .4%; minor mode = .5 ± .2%; minor mode with distribution instructions .7 ± .3%; $F(2, 190) = .80$, $p = .37$; no interaction: $F(1, 190) = .28$, $p = .60$; Figure 4D). Similarly, the average spatial error on the explicit paper test increased over time (main effect of retention interval: 15m = −.54 ± .08 AU; 28d = −.23 ± .07 AU; $F(1, 190) = 8.83$, $p = .003$; no main effect of learning condition: major mode = −.32 ± .09 AU; minor mode = −.41 ± .06 AU; minor mode with distribution instructions −.52 ± .09 AU; $F(2, 190) = .000$, $p = .99$; no interaction: $F(1, 190) = .20$, $p = .66$; Figure 4E).

On the navigation test, we found better gist memory (lower KL divergence) after 15 min than 28 days (main effect of retention interval: 15m = −.03 ± .008; 28d = −.005 ± .005; $F(1, 190) = 9.52$, $p = .002$; Figure 4B). There was no main effect of learning condition (major mode = −.01 ± .008; minor mode = −.02 ± .006; minor mode with distribution instructions −.03 ± .008; $F(2, 190) = .37$, $p = .54$), nor an interaction between retention interval and learning condition, $F(1, 190) = 1.14$, $p = .29$. On the paper test, we similarly found better gist memory after 15 min than 28 days (main effect of retention interval: 15m = −.10 ± .01; 28d = −.03 ± .01; $F(1, 190) = 10.2$, $p = .002$) but no main effect of learning condition (major mode = −.05 ± .02; minor mode = −.07 ± .01; minor mode with distribution instructions −.08 ± .02; $F(2, 190) = .02$, $p = .90$) or interaction, $F(1, 190) = 1.7$, $p = .19$ (Figure 4C). Therefore, we found no evidence that number of distributions, learning order, or instructions were responsible for the results of the earlier experiments. In summary, Experiment 3 supported evidence from prior experiments that both individual and gist memory decreased over time but did not support the idea that retention memory is biased toward the last learned location or that offering different test instructions emphasizing the distributed nature of locations changed performance.

General Discussion

Across three experiments, we found that gist representations of spatially structured experiences emerged during learning and decreased over time. This occurred (a) in a similar manner to memory for individually learned locations, (b) whether or not the mean location overlapped with these locations, and (c) whether there were one or two learned distributions. These findings differ from an account proposing that structured knowledge improves over time (Richards et al., 2014) and with our predictions that such gist would take time to emerge but would also decline at long delays (i.e., maximal at intermediate retention intervals).

In category learning, false memory, and schema-related paradigms, generalization from learned information to centrally related information occurs immediately (Bower et al., 1979; Posner & Keele, 1968; Read, 1996; Roediger & McDermott, 1995; Spencer & Hund, 2002). Similarly, here we found evidence of such generalization at early time intervals, which was generally maintained for at least one day before gradually decreasing (see Tompary et al., 2020, Zeng et al., 2021, and Berens et al., 2020, for very recent converging results). Our findings support recent complementary learning systems models (McClelland, 2013; Schapiro et al., 2017) proposing that generalization can occur rapidly. The idea that gist forms rapidly also supports fuzzy trace theory, which posits that detailed and gist representations are created in parallel (Reyna & Brainerd, 1995). This theory additionally states that gist representations do not become stronger over time, concurring with our results.

Another fitting psychological model is the category adjustment model, which suggests that detailed spatial memories are nested within categories and memory for them become adjusted based on category means (Huttenlocher et al., 1991). Here, coin and diamond labels make up the categories and the individual object locations make up the exemplars, and adjusting responses toward the mean of these categories could explain findings such as the false memories seen in Experiment 2. Although our results could seem at odds with the finding that category adjustment is greater at longer lags, these lags in prior research are either less than a minute in length (Holden et al., 2013) or years long with repeated intervening exposure, such as for spatial memory within a college campus (Uttal et al., 2010). Therefore, it could be that the likelihood of making category adjustments decreases over longer periods of time without repeated exposure.

It is important to point out that the discrepancy between our findings and Richards et al. (2014) could be accounted for by methodological differences between the studies. First, there are clear differences in cognitive capacity between rodents and humans and, relatedly, in the salience of swimming to safety versus navigating using only forward movements and rotations in virtual reality. Second, in Richards et al. (2014), trials were fixed in a particular order for every animal, whereas in our study this order was randomized for each subject. We explored whether the order mattered in Experiment 3, in particular whether there was a recency bias, by manipulating the location on the final trial. However, we did not find evidence for a recency effect bias, leaving inconclusive whether trial order could explain the differing dynamics of consolidation. Third, in Richards et al. (2014), single platform locations were repeated four times on each day, whereas here each location was learned only once. The lack of repetitions here may limit the result by reducing the strength of individual memories and reduce memory representations after longer delays. Finally, and perhaps most importantly, in Richards et al. (2014) learning occurred across several days prior to the retention interval, whereas in our study, learning occurred in a single session. Interestingly, two other recent papers using single-session learning similarly found that gist-like representations slowly faded over time (Berens et al., 2020; Tompary et al., 2020). However, spaced learning (e.g., Cepeda et al., 2008)—once proposed to be "the enemy of induction" (see Kornell & Bjork, 2008, p. 585)—has been shown to promote generalization via abstraction of relevant features and forgetting of irrelevant features (Vlach, 2014; Vlach & Sandhofer, 2012; Vlach et al., 2008). It remains an open question as to whether the rapidly formed, gist-like representations shown here differ from those formed in a slow, spaced fashion, and allowing time to elapse between related experiences is an important consideration for future research.

There are some other important limitations to these studies. Although individual location memory was above chance at the earliest retention intervals, we did not ask for specific memories (e.g., "where was the fifth learned coin location?") in either the navigation or paper test, limiting the precision with which individual memory could be ascertained. Experiments 2 and 3 had smaller object sizes by design to accommodate having a distribution with a full, nonexistent platform in its mean location. This led to smaller learning effect sizes from first to last learning trials in experiment 2 (dz = .18 for coins, dz = .23 for diamonds) and Experiment 3 (dz = .22 for the bimodal distribution) relative to Experiment 1 (dz = .47 for coins, dz = .54 for diamonds). This also led, especially in Experiment 2, to a high exclusion rate due to subjects not finishing learning within the allotted time. This was an a priori exclusion criterion and did not differ by retention interval condition, but it may limit the generalizability of these findings to the high-performing subject population. Furthermore, in Experiment 2, more females were excluded than males, leading to a possible limitation of how the results may generalize. Regarding gender differences, however, we found inconclusive supporting evidence from exclusion numbers in Experiments 1 and 3.

Last, we may be limited in our ability to interpolate possible differences within the intermediate retention intervals. In Experiment 3, we prioritized manipulating instructions at test and the location of the last learning trial at the earliest and latest retention intervals (leading to six total conditions) rather than testing all retention intervals from Experiments 1 and 2 (four total conditions). We made this decision because Experiments 1 and 2 showed no U-shaped results over time as we had initially predicted. Finally, there could be effects at earlier ($<$ 15 min) or longer ($>$ 28 d) intervals that could prove important to the story.

One interesting feature of our results in Experiment 2 is that we found trends toward decreased relative evidence for the mean representation compared with the learned locations over time. To the extent that gist is represented by a true statistical average "blending" of experiences, this measure shows evidence against even the relative strength of gist increasing over time. However, gist could also represent a distribution of the learned locations—with less precise detail, but a distribution nonetheless—without necessarily representing a blended average across experiences. Yet another possibility is that we see this trend because of relatively weakly learned individual experiences of finding platforms only a single time. In the DRM paradigm, false memory negatively correlates with veridical memory (Roediger et al., 2001), and therefore the effects of gist remaining better maintained over time could result more from a loss of veridical memories than a strengthening of gist, per se. Here, with arguably weaker initial true memories than in other studies, these relative stabilities could differ. We also note that, although the quantitative nature of these tendencies differs from prior studies, in Experiment 2 they were only trends; as such, we hope future work disambiguates these possibilities.

Ultimately, our findings are more ambivalent to the idea that off-line processes, such as hippocampal replay of these learned experiences (Richards et al., 2014; Stickgold & Walker, 2013), continue to contribute in substantial ways to memory generalization. Under different conditions, such as with different paradigms, with spaced learning, or with stronger learning, off-line processes may enrich these representations. However, here, with gist representations declining over time in every experiment, it appears that forgetting overwhelms any off-line processes that positively contribute to generalization.

References

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

Berens, S. C., Richards, B. A., & Horner, A. J. (2020). Dissociating memory accessibility and precision in forgetting. *Nature Human Behaviour*, *4*(8), 866–877. https://doi.org/10.1038/s41562-020-0888-8

Bower, G. H., Black, J. B., & Turner, T. J. (1979). Scripts in memory for text. *Cognitive Psychology*, *11*(2), 177–220. https://doi.org/10.1016/0010-0285(79)90009-4

Cepeda, N. J., Vul, E., Rohrer, D., Wixted, J. T., & Pashler, H. (2008). Spacing effects in learning. *Psychological Science*, *19*(11), 1095–1102. https://doi.org/10.1111/j.1467-9280.2008.02209.x

Fiser, J., & Aslin, R. N. (2001). Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychological Science*, *12*(6), 499–504. https://doi.org/10.1111/1467-9280.00392

Gallo, D. A. (2010). False memories and fantastic beliefs: 15 years of the DRM illusion. *Memory & Cognition*, *38*(7), 833–848. https://doi.org/10.3758/MC.38.7.833

Gekas, N., Chalk, M., Seitz, A. R., & Seriès, P. (2013). Complexity and specificity of experimentally-induced expectations in motion perception. *Journal of Vision*, *13*(4), 8. https://doi.org/10.1167/13.4.8

Graves, K. N., Antony, J. W., & Turk-Browne, N. B. (2020). Finding the pattern: Online extraction of spatial structure during virtual navigation. *Psychological Science*, *31*(9), 1183–1190.

Holden, M. P., Newcombe, N. S., & Shipley, T. F. (2013). Location memory in the real world: Category adjustment effects in 3-dimensional space. *Cognition*, *128*(1), 45–55. https://doi.org/10.1016/j.cognition.2013.02.016

Huttenlocher, J., Hedges, L. V., & Duncan, S. (1991). Categories and particulars: Prototype effects in estimating spatial location. *Psychological Review*, *98*(3), 352–376. https://doi.org/10.1037/0033-295X.98.3.352

Kornell, N., & Bjork, R. A. (2008). Learning concepts and categories. *Psychological Science*, *19*(6), 585–592. https://doi.org/10.1111/j.1467-9280.2008.02127.x

McClelland, J. L. (2013). Incorporating rapid neocortical learning of new schema-consistent information into complementary learning systems theory. *Journal of Experimental Psychology: General*, *142*(4), 1190–1210. https://doi.org/10.1037/a0033812

McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). 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), 419–457. https://doi.org/10.1037/0033-295X.102.3.419

Morris, R. (1984). Developments of a water-maze procedure for studying spatial learning in the rat. *Journal of Neuroscience Methods*, *11*(1), 47–60. https://doi.org/10.1016/0165-0270(84)90007-4

Nadel, L., & Moscovitch, M. (1997). Memory consolidation, retrograde amnesia and the hippocampal complex. *Current Opinion in Neurobiology*, *7*(2), 217–227. https://doi.org/10.1016/S0959-4388(97)80010-4

Podell, H. A. (1958). Two process of concept formation. *Psychological Monographs*, *72*(15), 1–20. https://doi.org/10.1037/h0093820

Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology*, *77*(3), 353–363. https://doi.org/10.1037/h0025953

Posner, M. I., & Keele, S. W. (1970). Retention of abstract ideas. *Journal of Experimental Psychology*, *83*(2, Pt. 1), 304–308. https://doi.org/10.1037/h0028558

Read, J. D. (1996). From a passing thought to a false memory in 2 minutes: Confusing real and illusory events. *Psychonomic Bulletin & Review*, *3*(1), 105–111. https://doi.org/10.3758/BF03210749

Reyna, V. F., & Brainerd, C. J. (1995). Fuzzy-trace theory: An interim synthesis. *Learning and Individual Differences*, *7*(1), 1–75. https://doi.org/10.1016/1041-6080(95)90031-4

Richards, B. A., Xia, F., Santoro, A., Husse, J., Woodin, M. A., Josselyn, S. A., & Frankland, P. W. (2014). Patterns across multiple memories are identified over time. *Nature Neuroscience*, *17*(7), 981–986. https://doi.org/10.1038/nn.3736

Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(4), 803–814.

Roediger, H. L., Watson, J. M., McDermott, K. B., & Gallo, D. A. (2001). Factors that determine false recall: A multiple regression analysis. *Psychonomic Bulletin & Review*, *8*(3), 385–407. https://doi.org/10.3758/bf03196177

Rosenbaum, R. S., Priselac, S., Köhler, S., Black, S. E., Gao, F., Nadel, L., & Moscovitch, M. (2000). Remote spatial memory in an amnesic person with extensive bilateral hippocampal lesions. *Nature Neuroscience*, *3*(10), 1044–1048. https://doi.org/10.1038/79867

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926–1928. https://doi.org/10.1126/science.274.5294.1926

Schapiro, A. C., Gregory, E., Landau, B., McCloskey, M., & Turk-Browne, N. B. (2014). The necessity of the medial temporal lobe for statistical learning. *Journal of Cognitive Neuroscience*, *26*(8), 1736–1747. https://doi.org/10.1162/jocn_a_00578

Schapiro, A. C., Kustner, L. V., & Turk-Browne, N. B. (2012). Shaping of object representations in the human medial temporal lobe based on temporal regularities. *Current Biology*, *22*(17), 1622–1627. https://doi.org/10.1016/j.cub.2012.06.056

Schapiro, A. C., Turk-Browne, N. B., Botvinick, M. M., & Norman, K. A. (2017). Complementary learning systems within the hippocampus: a neural network modelling approach to reconciling episodic memory with statistical learning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1711), 20160049. https://doi.org/10.1098/rstb.2016.0049

Smith, J. D., & Minda, J. P. (2002). Distinguishing prototype-based and exemplar-based processes in dot-pattern category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(4), 800–811.

Spencer, J. P., & Hund, A. M. (2002). Prototypes and particulars: Geometric and experience-dependent spatial categories. *Journal of Experimental Psychology: General*, *131*(1), 16–37. https://doi.org/10.1037/0096-3445.131.1.16

Squire, L. R., Genzel, L., Wixted, J. T., & Morris, R. G. (2015). Memory consolidation. *Cold Spring Harbor Perspectives in Biology*, *7*(8), a021766https://doi.org/10.1101/cshperspect.a021766

Stickgold, R., & Walker, M. P. (2013). Sleep-dependent memory triage: Evolving generalization through selective processing. *Nature Neuroscience*, *16*(2), 139–145. https://doi.org/10.1038/nn.3303

Strange, W., Keeney, T., Kessel, F. S., & Jenkins, J. J. (1970). Abstraction over time of prototypes from distortions of random dot patterns: A replication. *Journal of Experimental Psychology*, *83*(3, Pt. 1), 508–510. https://doi.org/10.1037/h0028846

Thapar, A., & McDermott, K. B. (2001). False recall and false recognition induced by presentation of associated words: Effects of retention interval and level of processing. *Memory & Cognition*, *29*(3), 424–432. https://doi.org/10.3758/bf03196393

Tompary, A., Zhou, W., & Davachi, L. (2020). Schematic memories develop quickly, but are not expressed unless necessary. *Scientific Reports*, *10*(1), 1–17.

Turk-Browne, N. B., Jungé, J. A., & Scholl, B. J. (2005). The automaticity of visual statistical learning. *Journal of Experimental Psychology: General*, *134*(4), 552–564. https://doi.org/10.1037/0096-3445.134.4.552

Uttal, D. H., Friedman, A., Hand, L. L., & Warren, C. (2010). Learning fine-grained and category information in navigable real-world space. *Memory & Cognition*, *38*(8), 1026–1040. https://doi.org/10.3758/MC.38.8.1026

Vlach, H. A. (2014). The spacing effect in children's generalization of knowledge: Allowing children time to forget promotes their ability to learn. *Child Development Perspectives*, *8*(3), 163–168. https://doi.org/10.1111/cdep.12079

Vlach, H. A., & Sandhofer, C. M. (2012). Distributing learning over time: The spacing effect in children's acquisition and generalization of science concepts. *Child Development*, *83*(4), 1137–1144. https://doi.org/10.1111/j.1467-8624.2012.01781.x

Vlach, H. A., Sandhofer, C. M., & Kornell, N. (2008). The spacing effect in children's memory and category induction. *Cognition*, *109*(1), 163–167. https://doi.org/10.1016/j.cognition.2008.07.013

Wiltgen, B. J., & Silva, A. J. (2007). Memory for context becomes less specific with time. *Learning & Memory*, *14*(4), 313–317. https://doi.org/10.1101/lm.430907

Winocur, G., Moscovitch, M., & Bontempi, B. (2010). Memory formation and long-term retention in humans and animals: Convergence towards a transformation account of hippocampal-neocortical interactions. *Neuropsychologia*, *48*(8), 2339–2356. https://doi.org/10.1016/j.neuropsychologia.2010.04.016

Winocur, G., Moscovitch, M., & Sekeres, M. (2007). Memory consolidation or transformation: Context manipulation and hippocampal representations of memory. *Nature Neuroscience*, *10*(5), 555–557. https://doi.org/10.1038/nn1880

Woolley, D. G., Laeremans, A., Gantois, I., Mantini, D., Vermaercke, B., Op de Beeck, H. P., Swinnen, S. P., Wenderoth, N., Arckens, L., & D'Hooge, R. (2013). Homologous involvement of striatum and prefrontal cortex in rodent and human water maze learning. *Proceedings of the*

National Academy of Sciences of the United States of America, *110*(8), 3131–3136. https://doi.org/10.1073/pnas.1217832110

Woolley, D. G., Vermaercke, B., Op de Beeck, H., Wagemans, J., Gantois, I., D'Hooge, R., Swinnen, S. P., & Wenderoth, N. (2010). Sex differences in human virtual water maze performance: Novel measures reveal the relative contribution of directional responding and spatial knowledge. *Behavioural Brain Research*, *208*(2), 408–414. https://doi.org/10.1016/j.bbr.2009.12.019

Zeng, T., Tompary, A., Schapiro, A. C., & Thompson-Schill, S. (2021). Tracking the relation between gist and item memory over the course of long term memory consolidation. *bioRxiv*. https://www.biorxiv.org/content/10.1101/2021.01.05.425378v1

Received May 5, 2020
Revision received February 24, 2021
Accepted March 1, 2021 ■