Overlap among spatial memories triggers divergence of hippocampal representations

Avi J. H. Chanales[1], Ashima Oza[1], Serra E. Favila[1], Brice A. Kuhl[2]

[1]Department of Psychology, New York University, 6 Washington Pl., New York, New York 10003, USA

[2]Department of Psychology and Institute of Neuroscience, 1227 University of Oregon, Eugene, Oregon 97403, USA

ABSTRACT

Across the domains of spatial navigation and episodic memory, the hippocampus is thought to play a critical role in forming distinct representations of overlapping events. However, it is less clear how and when hippocampal representations of overlapping events become distinct. Here, using spatiotemporal pattern analysis of human fMRI data, we measured learning-related changes in hippocampal representations of overlapping real-world routes. We found that learning resulted in highly targeted representational changes within the hippocampus that specifically reduced similarity among overlapping routes. Strikingly—and in contrast to representations in other navigation-related brain regions—hippocampal representations of overlapping routes diverged to the point where overlapping routes became less similar than non-overlapping routes. This learning-related plasticity preferentially occurred for hippocampal voxels that were initially shared across overlapping routes. Collectively, these findings indicate that overlap among event representations triggers a divergence of hippocampal activity patterns that dramatically reshapes representational structure.

INTRODUCTION

Distinct experiences often contain overlapping elements, creating the potential for memory interference. For example, a single location (e.g., a living room) may be the site of many different experiences. The hippocampus is widely thought to play a critical role in coding overlapping events such that interference is minimized. Compelling evidence for this function comes from intracranial recordings in rodents during spatial navigation. For example, when rodents alternate between left- and right-hand turns in a T-maze, cells within the hippocampus differentially fire during the central stem (the overlapping path), according to whether the current route is a 'right-turn' or 'left-turn' route[1,2]. Similarly, human fMRI evidence indicates that elements which are shared across multiple event sequences are distinctly coded in the hippocampus according to the specific sequence to which they belong 3. While these studies and others have led to general agreement that the hippocampus forms distinct codes for overlapping experiences[4-14], it is less clear how and when hippocampal codes become distinct.

One way the hippocampus may minimize interference between overlapping events is by orthogonalizing, or pattern separating, neural activity patterns[15-23]. For example, if two spatially-overlapping routes are orthogonally coded, their neural representations should be no more similar than those corresponding to non-overlapping routes. A more extreme, and seemingly paradoxical, outcome is that hippocampal representations of overlapping events may be even less similar than representations of non-overlapping events. Why might this occur? When events are highly overlapping, this may create competition between event representations that the hippocampus solves in a targeted way by pushing representations apart from one another. By analogy, when two children are bickering, a parent's reaction may be to move the children to opposite corners of the room. Thus, it may be that initial overlap among hippocampal representations triggers an 'over-compensation' wherein overlapping events actually become less similar than non-overlapping events[4]. The idea that overlap triggers a repulsion of event representations has been

referred to as 'differentiation' and is supported and inspired by computational models[24,25] and a limited number of empirical observations [4,26–29].

Importantly, hippocampal differentiation is thought to occur gradually, as a result of repeatedly experiencing or remembering an event [26]. This can be contrasted with the idea of pattern separation, where sparse coding in the hippocampus gives rise to an immediate and automatic orthoganilzation of activity patterns [30]. Moreover, because differentiation is thought to be triggered by competition, neural activity patterns should change in a particular way. Namely, differentiation should preferentially occur for features within activity patterns that are initially shared across event representations. Indeed, differentiation can be thought of as an un-sharing of initially-shared features.

Here, we used a real-world, human spatial-learning task inspired by canonical rodent T-Maze paradigms to test for hippocampal differentiation of overlapping routes. Using spatiotemporal analysis of fMRI data, we tracked learning-related changes in the similarity of hippocampal representations of overlapping and non-overlapping routes. We predicted route overlap would trigger a repulsion of hippocampal representations, with similarity among overlapping routes dropping below similarity among non-overlapping routes[4]. Critically, while such a result would be readily explained by the differentiation account, it would not be explained by a pattern separation account. Additionally, to address the specific way in which hippocampal activity patterns changed with learning, we tested whether learning-related differentiation was related to the degree to which individual hippocampal voxels were initially shared across overlapping events.

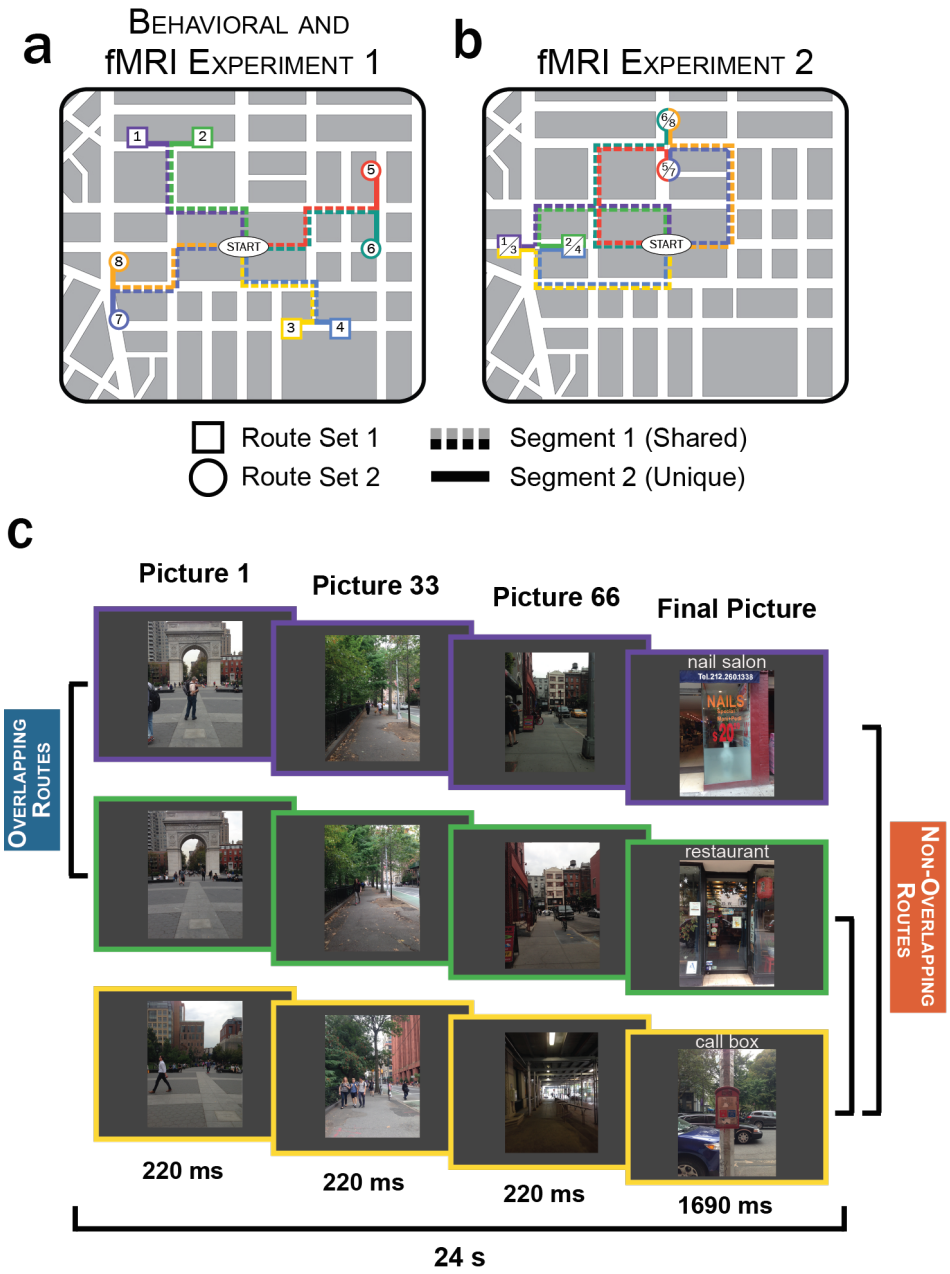


Figure 1. Route stimuli and experimental design. (**a**) In the behavioral experiment and fMRI Experiment 1, stimuli consisted of 8 routes that traversed the New York University campus. Each subject learned 4 routes–either Set 1 (routes 1-4) or Set 2 (routes 5-8). Each set included pairs of routes that shared a common path ('overlapping routes'; e.g. routes 1 and 2) and pairs of routes with no common paths ('non-overlapping routes'; e.g. routes 1 and 3). Individual routes contained two segments: Segment 1 refers to the portion of each route that shared a path with another route; Segment 2 refers to the unique portion of each route (after the overlapping routes diverged). (**b**) In fMRI Experiment 2 the stimuli again consisted of 8 routes with each subject learning 4 of the 8 routes. However, in addition to overlapping and non-overlapping route pairs, the routes in Experiment 2 also contained pairs that followed distinct paths but terminated at the same location ('same destination'; e.g. routes 1 and 3). (**c**) In each experiment, each route learning trial lasted a total of 24s and was comprised of a sequence of rapidly presented pictures.

RESULTS

Behavioral measures of route discrimination

In an initial behavioral experiment subjects studied sets of real-world routes that traversed the New York University campus. For each subject, the set of routes included (a) pairs that shared a common path before diverging to terminate at distinct destinations ('overlapping routes') and (b) pairs with no paths in common ('non-overlapping routes') (Fig. 1a). Importantly, each route contributed to both conditions. For example, 'route 1' and 'route 2' were overlapping routes, but 'route 1' and 'route 3' were non-overlapping routes (Fig. 1c). Each route contained an initial segment that was shared with another route (Segment 1), and a later segment, including the destination, that was route-specific (Segment 2; Fig. 1a). Although the real-world spatial locations of the overlapping segments were identical, the pictures for each route were taken at different times and therefore differed subtly in terms of pedestrians, vehicles, etc. (Fig. 1c and Supplementary Videos 1-8). Routes were studied twice per round for 14 rounds. Subjects were instructed to learn each route (i.e., the specific path to each destination) but were not told the destination at the start of the route. After each study round, subjects were shown individual pictures drawn from the routes and were asked to select the destination associated with each picture. Of central interest was accuracy for pictures drawn from Segment 1 of each route because selecting the correct destination for these pictures required discriminating between overlapping routes. Overall, subjects selected the correct destination ('target') at a higher rate than the destination of the overlapping route ('competitor') ($F_{1,21}$ = 43.31, P = 0.000002; Fig. 2a). The relative rate of target vs. competitor responses also markedly increased across learning rounds ($F_{1,21}$ = 38.11, P = 0.000004; Fig. 2b,c).

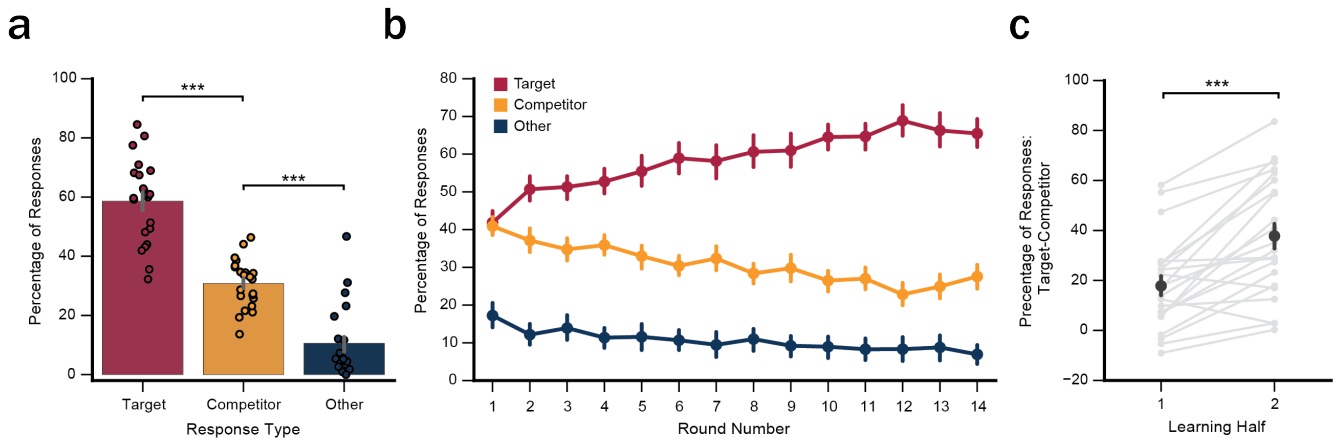


Figure 2. Memory performance for Segment 1 pictures in the behavioral study. (**a**) After each learning round subjects were shown static images sampled from each route and were asked to choose the corresponding destination from a set of four picture options: the target destination, the destination associated with the overlapping route ('competitor') or two destinations associated with the non-overlapping routes ('other'). Subjects were significantly more likely to select the target destination than competitor destination ($F_{1,21}$ = 43.31, P = 0.000002) and significantly more likely to choose the competitor destination than other destinations ($F_{1,21}$ = 41.39, P = 0.000002), despite the fact that 'other' options were more prevalent (2/4) than competitor options (1/4). (**b**) The relative percentage of target vs. competitor responses markedly increased over learning rounds ($F_{1,21}$ = 38.11, P = 0.000004). (**c**) Discrimination between overlapping routes (percentage target responses - competitor responses) was significantly greater in the 2nd half of learning than the 1st half (t_{21} = 5.78, P = 0.00001). Error bars reflect +/- SEM. Colored dots reflect data from individual subjects. *** $P <$ 0.001

Hippocampal representations of overlapping routes diverge with learning

We next tested for hippocampal differentiation of overlapping routes in two fMRI studies. The first fMRI study used the same stimuli as the behavioral study (Fig. 1a). The second fMRI study used a new set of stimuli that again included overlapping and non-overlapping routes, but some of the non-overlapping routes terminated at a common destination (Fig. 1b). Unless otherwise noted, all analyses below combine data across experiments and all comparisons of non-overlapping routes are restricted to those that terminated at distinct destinations.

For Segment 1 of each route, we obtained a corresponding neural activity pattern by extracting voxel-wise patterns of activity as they unfolded, over time. These spatiotemporal activity patterns were then correlated for every pair of routes, resulting in a correlation matrix reflecting pairwise route similarity (Fig. 3a). We

considered pattern similarity for (1) repetitions of the same route, (2) overlapping routes, and (3) non-overlapping routes. Separate correlation matrices were generated for each subject's hippocampus and for a control region: the 'parahippocampal place area' (PPA), which is adjacent to the hippocampus and is involved in scene processing and navigation[31,32] (Fig. 3b). Because our behavioral experiment indicated that discrimination of overlapping routes robustly improved from the 1st to 2nd half of learning (Fig. 2c), we divided the fMRI data into halves and independently computed pattern similarity measures within each of these halves. As in the behavioral experiment, subjects in both fMRI experiments were able to successfully discriminate between the overlapping routes by the end of learning (Supplementary Fig. 1).

Of critical interest, there was a learning-related decrease in pattern similarity among overlapping compared to non-overlapping routes, as reflected by an interaction between overlap (overlapping/non-overlapping) and learning (1st half/2nd half) ($F_{1,39}$ = 13.163, P = 0.0008; Fig. 3c). Whereas pattern similarity among overlapping routes decreased with learning ($F_{1,39}$ = 35.21, P = 0.0000006), similarity among non-overlapping routes did not change ($F_{1,39}$ = 0.24, P = 0.63; Fig. 3e). This dissociation is striking when considering that all routes contributed to both the overlapping and non-overlapping comparisons. Thus, learning did not globally reduce similarity among routes; rather, learning specifically reduced similarity between overlapping routes. Moreover, overlapping route similarity decreased to the point that in the 2nd half of learning overlapping routes were markedly less similar than non-overlapping routes ($F_{1,39}$ = 14.20, P = 0.0005; Fig. 3f). This result was significant in each of the fMRI experiments (Ps < .05; see Supplementary Fig. 2 for results separated by experiment and Supplementary Fig. 3 for results separated by stimulus set). Thus, despite the fact that overlapping routes were spatially and visually more similar than non-overlapping routes, the hippocampus represented overlapping routes as less similar than non-overlapping routes–a result we refer to as a 'reversal effect' because it is opposite to the inherent similarity structure. This reversal effect was not present in the 1st half of learning ($F_{1,39}$ = 1.41, P = 0.24),

confirming that it was a result of learning (See Supplementary Fig. 4 for hippocampal pattern similarity computed at a finer time scale). Moreover, when considering data from Segment 2–i.e., after overlapping routes diverged–the reversal effect was absent (see Supplementary Fig. 5). Thus, hippocampal activity patterns became relatively more similar precisely when the overlapping routes diverged. These data indicate that divergence of hippocampal activity patterns was a direct response to overlap.

Within PPA, there was no learning-related reduction in the similarity of overlapping compared to non-overlapping routes ($F_{1,39}$ = 2.42, P = 0.13; Fig. 3d**).** In fact, overlapping route similarity was greater than non-overlapping route similarity in the 1st half ($F_{1,39}$ = 21.01, P = 0.00005) and 2nd half of learning ($F_{1,39}$ = 4.63, P = 0.038; Fig. 3f; note: this effect differed across Experiments, see Supplementary Fig. 6). Thus, the inverted representational structure that was observed in hippocampus by the end of learning was absent in PPA. The dissociation between representational structure in PPA vs. hippocampus at the end of learning was reflected in a highly significant region x overlap interaction ($F_{1,39}$ = 22.18, P = 0.00003). A similar dissociation was observed when comparing hippocampus to retrosplenial cortex, another region involved in scene-processing and navigation (see Supplementary Fig. 7).

Although we primarily focus on the comparison of overlapping vs. non-overlapping routes, the comparison of overlapping vs. same routes is also informative. To the extent that overlapping route representations are distinct, overlapping route similarity should be lower than same route similarity. Indeed, within the hippocampus there was a learning-related change such that overlapping route similarity decreased relative to same route similarity ($F_{1,39}$ = 7.59, P = 0.009). Overlapping route similarity was significantly lower than same route similarity in the 2nd half of learning ($F_{1,39}$ = 5.61, P = 0.023), but not in the 1st half of learning ($F_{1,39}$ = 0.85, P = 0.35). Within PPA, there was no learning-related change in overlapping vs. same route similarity ($F_{1,39}$ = 0.003, P = 0.96). The difference

between overlapping and same route similarity was not significant in the 1st half ($F_{1,39}$ = 0.89, P = 0.35) or 2nd half ($F_{1,39}$ = 0.86, P = 0.36).

One way in which hippocampal route representations may diverge with learning is through the learned ability to predict the route destinations[33-38]. To test this possibility we considered data from Experiment 2, which contained pairs of non-overlapping routes that terminated at distinct destinations as well as pairs of non-overlapping routes that terminated at the same destination. If hippocampal activity patterns reflected navigational goals, spatiotemporal pattern similarity from Segment 1 should be greater for non-overlapping routes that terminated at the same destination relative to non-overlapping routes that terminated at distinct destinations. However, we found no evidence for a difference between these conditions (Supplementary Fig. 8). Specifically, there was no learning related increase in spatiotemporal similarity for same destination pairs relative to distinct destination pairs ($F_{1,20}$ = 0.53, P = 0.47), nor was there a difference between same and distinct destination pairs when considering second-half data alone (t_{20} = 0.98, P = 0.34). Thus, we did not observe any evidence to suggest that destination coding was contributing to the divergence of hippocampal activity patterns.

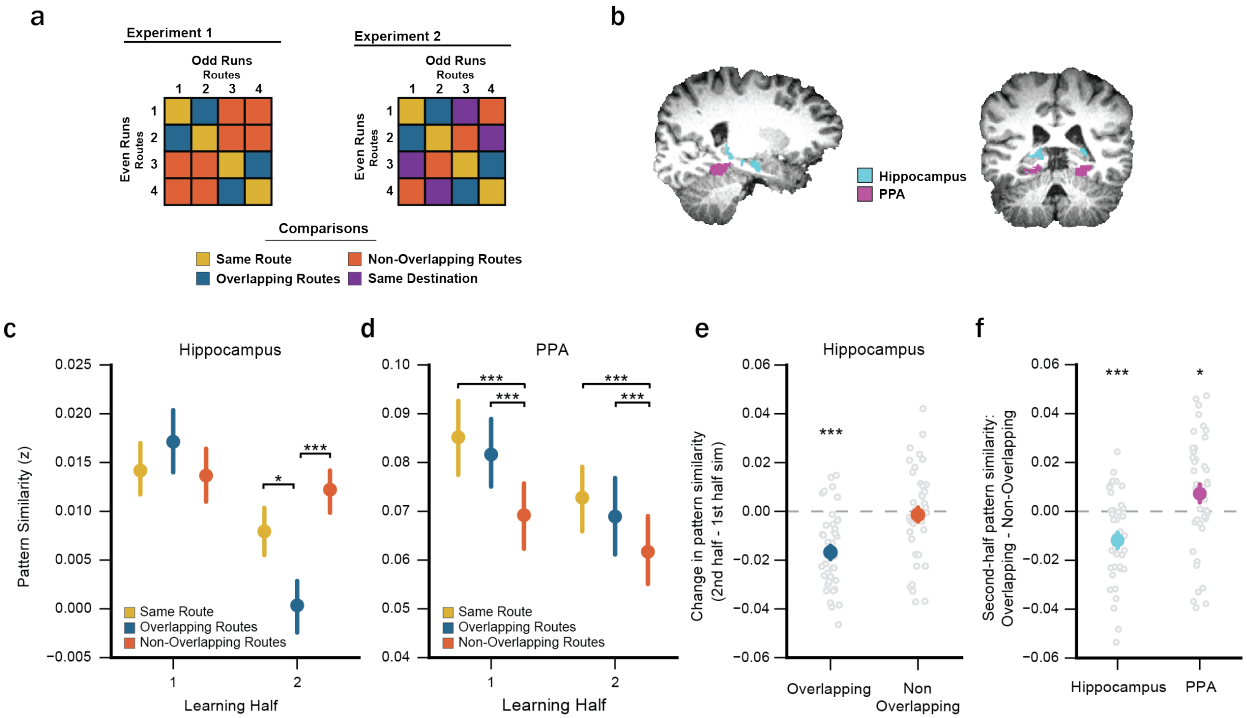


Figure 3. Learning-related changes in hippocampal pattern similarity. (a) Sample similarity matrices depicting analyses for Experiments 1 and 2. For each experiment, pearson correlations were applied to spatiotemporal activity patterns to measure the similarity between: repetitions of the same route ('same route'), routes with overlapping paths but distinct destinations ('overlapping routes'), and routes with non-overlapping paths and distinct destinations ('non-overlapping routes'). Experiment 2 included an additional comparison of routes with non-overlapping paths that ended at a common destination ('same destination'), which is described in Supplementary Fig. 6. All correlations were applied to spatiotemporal activity patterns from independent fMRI runs (odd vs. even runs). **(b)** Hippocampus and parrahipocampal place area (PPA) regions of interest for a representative subject, displayed on T1 anatomical scan. **(c)** Within the hippocampus, learning dramatically reduced the relative similarity of overlapping routes as reflected by learning-related changes (1st vs. 2nd half) in the comparison between same vs. overlapping route similarity (P = 0.009) and in the comparison between overlapping vs. non-overlapping route similarity (P = 0.0008). **(d)** Within PPA, there was no learning-related change in the comparison between same vs. overlapping route similarity (P = 0.96) or in the comparison between overlapping vs. non-overlapping route similarity (P = 0.13) **(e)** Within the hippocampus, overlapping route similarity decreased across learning (1st half vs. 2nd half, P = 0.0000006) whereas non-overlapping route similarity did not change with learning (P = 0.63) **(f)** In the 2nd half of learning, overlapping route similarity was significantly lower than non-overlapping route similarity within the hippocampus (P = 0.0005; 'reversal effect') whereas in PPA the opposite was true: overlapping route similarity was significantly greater than non-overlapping route similarity (P = 0.038). Error bars reflect SEM. Light grey dots reflect data from individual subjects. * $P <$ 0.05, ** $P <$ 0.01, *** $P <$ 0.001.

Voxel-level plasticity is triggered by pattern overlap

The preceding results indicate that hippocampal representations of overlapping events diverged with learning. But how, exactly, were activity patterns shaped by learning? Putatively, differentiation of event representations is a reaction to competition[26,39], with a competing memory subject to plasticity (and corresponding differentiation) to the extent that it becomes active during the processing of a target memory. Interestingly, however, it has been argued that this relationship between activation and differentiation is non-monotonic. Namely, differentiation is most likely to occur when competing memories are 'moderately activated.' When a competing memory is moderately activated, it is subject to inhibition that weakens associations with the target memory via the 'drop out' of overlapping features. However, when a competing memory is strongly activated, its features resist inhibition and the competing memory remains strongly associated with the target memory. Finally, if a competing memory is only weakly activated, there is little opportunity or need for plasticity to occur. This theory, has been referred to as the non-monotonic plasticity hypothesis and has been elegantly described in computational terms[24,25] and can account for several empirical observations that are otherwise difficult to explain[4,26–28,40,41]. We reasoned that this theory may also explain voxel-level changes across learning that were observed in the present study. Namely, within an individual voxel, we can think of a pair of overlapping routes as corresponding to partially overlapping assemblies of neurons. If the assemblies are highly overlapping, then presentation or processing of one route is likely to activate the overlapping route, via spread of activation. Although we do not have a way to directly measure the overlap of neural ensembles within a voxel, we reasoned that if a voxel responds similarly to two overlapping routes, then it is likely that the two routes are represented by overlapping ensembles of neurons (within that voxel). In other words, the degree to which a voxel responds similarly to a pair of overlapping routes may provide a clue as to the overlap in the underlying neural representations. If so, then following the non-monotonic plasticity hypothesis, we

would expect differentiation to maximally occur for voxels that initially represented overlapping routes as 'moderately similar' (Fig. 4a).

To test this prediction, the critical first step was to quantify the similarity with which individual voxels responded to each pair of routes at the beginning of learning. To do so, we correlated the timecourse of each voxel's activity for each pair of routes, using Segment 1 data only (see Methods). For a given pair of routes, voxels with relatively high timecourse similarity were considered to be voxels that were more strongly 'shared' across the routes (Fig. 4b). Timecourse similarity values were first obtained based on data from the 1st half of learning. Voxels were then rank-ordered by 1st-half timecourse similarity and binned into groups corresponding to 'weak,' 'moderate,' or 'strong' values (i.e., the bottom 1/3, middle 1/3, and top 1/3 of similarity values). Importantly, this binning was independently repeated for every pair of routes, each region of interest, and each subject. Thus, a given voxel may have been strongly shared across one pair of routes but weakly shared across a different pair of routes. Timecourse similarity values from the 2nd half of learning were then obtained from these voxel bins, which allowed timecourse similarity values at the end of learning to be considered as a function of timecourse similarity at the beginning of learning. It is important to note that we did not measure changes in time course similarity from the first to second half, so as to avoid problems associated with regression to the mean.

Within the hippocampus, 2nd-half timecourse similarity for overlapping routes significantly varied according to 1st-half similarity ($F_{2,78}$ = 4.74, P = 0.012), with the lowest levels of 2nd-half similarity occurring for voxels that were 'moderately' shared during the 1st half of learning (Fig. 4c). Qualitatively, the relationship strongly matched the predicted non-monotonic pattern (Fig. 4a). For non-overlapping routes, 2nd-half timecourse similarity did not vary according to 1st half similarity ($F_{2,78}$ = 0.28, P = 0.76). An ANOVA with factors of overlap (overlapping/non-overlapping) and bin (weak/moderate/strong) revealed a significant overlap x bin interaction ($F_{2,78}$ = 3.19, P = 0.046), with the difference between overlapping and non-overlapping routes (the reversal effect) highly

significant in the 'moderate' bin ($F_{1,39}$ = 19.17, P = 0.00009), marginally significant in the 'weak' bin ($F_{1,39}$ = 3.62, P = 0.064), and not significant in the 'strong' bin ($F_{1,39}$ = 1.53, P = 0.22). Thus, consistent with predictions, voxels that were moderately shared across routes at the beginning of learning exhibited the strongest differentiation. The relationship between 1st and 2nd half timecourse similarity for overlapping routes was markedly different in PPA, as reflected by a significant region (hippocampus/PPA) x bin interaction ($F_{2,78}$ = 18.12, P = 0.0000003) and a marginally significant region x bin x overlap interaction ($F_{2,78}$ = 2.95, P = 0.0584). In PPA, patterns remained relatively stable as voxels that were moderately shared in the 1st half of learning remained moderately shared in the 2nd half of learning (Fig. 4d).

To more formally assess the non-monotonic relationship between 1st-half and 2nd-half timecourse similarity, we used a Bayesian curve-fitting algorithm–the Probablistic Curve Induction and Testing Tooblbox (P-CIT)[42]–that was specifically developed to test for non-monotonic relationships. Relative to quadratic trend analyses, the P-CIT algorithm has important advantages. Namely, the P-CIT algorithm allows for more detailed specification of a predicted curve shape, by explicitly including a set of curve parameters. In our case, the parameters describe the relationship between first-half timecourse similarity (x-axis) and second-half timecourse similarity (y-axis). Consistent with prior usage of the P-CIT algorithm 41, we parameterized the predicted curve shape as one in which the function, when moving from left to right, drops below the initial start value and then rises above the start value. The 'dip' in the middle of the curve reflects the prediction that differentiation (in the second-half of learning) should be greatest for voxels that were 'moderately' shared at the beginning of learning. The rise above the starting point after the 'dip' reflects the prediction that voxels that were strongly shared at the beginning of learning will remain (relatively) strongly shared at the end of learning. Using the specified curve parameters, the algorithm estimates a probability distribution over curves, conditional on the observed data. To assess evidence in favor of the predicted curve shape, the algorithm labels each sample

curve as theory consistent (in our case, if it drops below the starting value and then rises above the starting value) or inconsistent and then computes a log Bayes factor value that represents the log ratio of evidence in favor of or against the predicted shape. Positive log Bayes factor values indicate greater evidence in favor of the theory. For this analysis, we re-binned all of the 1st half timecourse similarity values into 60 bins (5 voxels per bin) in order to allow for greater variability in the observed curve shape. This analysis used data aggregated across all subjects [28,41,42].

The estimated curve (Fig. 4e) was consistent with the predicted curve shape (log Bayes factor = 1.51) and explained a significant amount of variance in the actual data (χ^2 = 11.13, P = 0.0008). We next ran a permutation test to estimate the null distribution of theory consistency values. Out of 500 permutations, only 2.6% yielded log Bayes factor values that matched or exceeded the value obtained from the un-permuted data, indicating that it was unlikely to obtain this level of support for the predicted curve shape by chance (Fig. 4f). Finally, to assesses the population-level reliability of the non-monotonic curve we ran a bootstrap resampling test in which we iteratively resampled data from subjects with replacement and then computed the log Bayes factor value for each iteration. Four-hundred and eighty-eight of the 500 bootstrap iterations (97.6%) yielded positive Bayes factor values. Thus, the curve-fitting analyses provided converging evidence for the predicted non-monotonic relationship between voxel overlap at the beginning vs. end of learning: hippocampal voxels that were 'moderately shared' across overlapping routes at the beginning of learning were the 'least shared' by the end of learning.

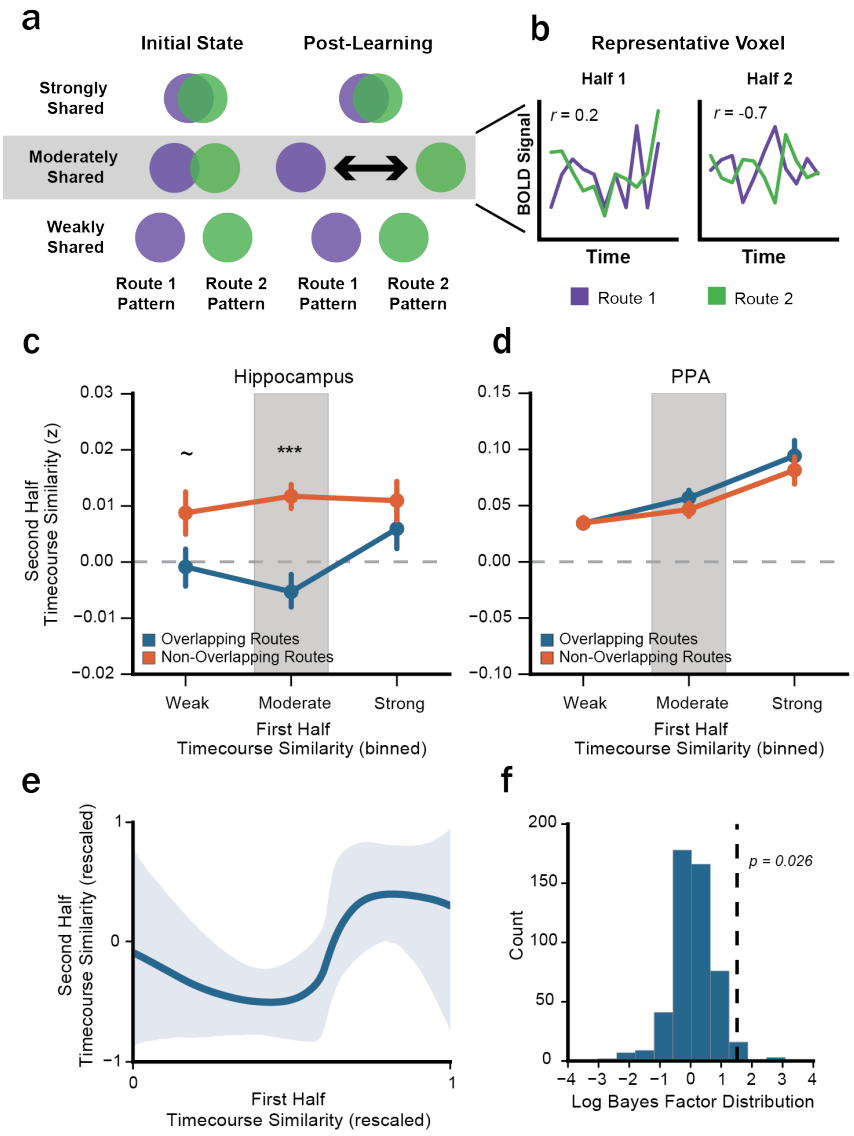


Figure 4. Voxel-level plasticity. (**a**) Schematic of non-monotonic plasticity. Each row corresponds to the position of a single voxel in representational space for each of two routes before and after learning. Overlap among circles reflects the degree to which a voxel is shared across route representations. According to the differentiation account, learning-related plasticity should maximally occur for voxels that start out as 'moderately shared' across representations, with the result of plasticity being a 'repulsion' of route representations. (**b**) Data from a single voxel illustrating a learning-related decrease in timecourse similarity. Timecourse similarity refers to the correlation of a single voxel's temporal pattern of activity across two difference routes. (**c,d**) Second-half timecourse similarity plotted as a function of first-half timecourse similarity, separately for the hippocampus (**c**) and PPA (**d**) and for overlapping (blue) and non-overlapping (orange) routes. Within the hippocampus, second-half timecourse similarity was markedly lower for overlapping than non-overlapping routes (a reversal effect) when voxels were moderately shared at the beginning of learning ($P = 0.00009$), consistent with the non-monotonic plasticity hypothesis. (**e**) Estimated plasticity curve relating 1st half timecourse similarity to 2nd half timecourse similarity for overlapping routes. Shaded area reflects the 90% credible interval. (**f**) Histogram showing distribution of log Bayes factor values–or the strength of evidence in favor of the non-monotonic hypothesis–generated from a permutation test. Only 2.6% of the 500 values from the permutation test were equal or greater to the observed value (dashed black line). Error bars reflect 95% confidence intervals. $P < 0.1$, *** $P < 0.001$.

DISCUSSION

Here, across two fMRI studies, we found that hippocampal representations of overlapping spatial routes dramatically diverged with learning–to the point that overlapping routes were coded as less similar than non-overlapping routes. This 'reversal effect' was clearly a result of learning as it was only evident after subjects gained considerable familiarity with the routes and it paralleled behavioral improvement in memory-based route discrimination. The result was also selective to the hippocampus, with no evidence of a reversal effect in PPA. Finally, using a novel analysis approach, we show that the learning-related divergence of hippocampal activity patterns was most likely to occur for voxels that were moderately shared across overlapping routes at the beginning of learning.

Several aspects of the stimuli we used and the analyses we employed are critical for the interpretation of our results. First, our analysis approach specifically compared representations of overlapping events to representations of non-overlapping events[4]. This allowed for learning-related changes to be expressed relative to a meaningful baseline. Indeed, the fact that hippocampal representations of visually- and spatially-overlapping routes became less similar than routes that contained no spatial overlap or visual similarity is not only striking, but it provides important insight into the underlying mechanism (a point we detail below). Second, our design did not involve separate sets of routes for the overlapping and non-overlapping comparisons; rather, each route was included in each 'condition.' For example, the comparison of routes 1 and 2 was a comparison of overlapping routes, whereas the comparison of routes 1 and 3 was a comparison of non-overlapping routes. As such, any observed differences between overlapping and non-overlapping routes cannot be attributed to differences between the actual stimuli or to differences in attention, familiarity, vigilance, etc. It is also of note that our findings generalized across different sets of stimuli (experiments 1 and 2). Lastly, for our critical comparison of overlapping vs. non-overlapping routes, we only considered spatiotemporal activity patterns during the overlapping segments of the routes (Segment 1 data)–that is, before the overlapping routes diverged. Indeed, once the

overlapping routes diverged (Segment 2 data), the hippocampal reversal effect 'disappeared' (Supplementary Fig. 5). Thus, hippocampal representations of overlapping routes were most dissimilar when routes actually overlapped, clearly indicating that the reversal effect was triggered by event overlap.

The mechanism that has most frequently been associated with disambiguating hippocampal representations is pattern separation[15-23]. However, the reversal effect we observed in the hippocampus is not readily explained by pattern separation. Rather, 'perfectly' pattern-separated representations of overlapping routes should have similarity values at–but not below–the similarity values for non-overlapping routes. In other words, pattern separation should only reduce global similarity among route representations[43]. It has also been argued that hippocampal activity patterns for overlapping events become disambiguated as event-specific contextual elements are learned and relational networks are formed[3,8]. Contextual elements may include event outcomes, locations, rewards, etc. However, because overlapping routes inherently shared more contextual features than non-overlapping routes, the observed reversal effect is difficult to explain based on a simple context-learning account. A related possibility is that hippocampal activity patterns reflected predictions about route destinations[33-38]. However, we did not observe any evidence in favor of destination coding (Supplementary Fig. 8) and, again, this account does not explain the reversal effect we observed. Rather, associating each route with a unique destination would only reduce global similarity among routes, similar to pattern separation or context learning. Thus, at least in the present study, hippocampal activity patterns appeared to reflect route information other than predicted destinations[13].

In contrast to accounts based on pattern separation, context learning, or destination coding, a differentiation account readily explains several critical aspects of our findings. First, according to a differentiation account, learning should result in highly-targeted representational changes where overlapping events are specifically distanced from one another. This prediction is supported by our observation of selective learning-related decreases among overlapping route representations in

the hippocampus. Second, if sufficient differentiation occurs, overlapping events should become less similar than non-overlapping events[4]. The observed reversal effect is therefore naturally explained by differentiation. Third, differentiation is thought to occur incrementally over the course of learning[26]. This can be contrasted with the idea that pattern separation reflects an automatic orthogonalization that occurs during initial event encoding owing to sparse coding within the hippocampus[16,30]. In other words, whereas differentiation is inherently a learning phenomenon, pattern separation is not. Our results clearly provide evidence of learning-related changes, consistent with the idea of differentiation. Indeed, the reversal effect only began to emerge after routes had been presented approximately 20 times (see Supplementary Fig. 4). Thus, differentiation occurred quite slowly. One question raised by the slow time scale of differentiation is why hippocampal representations continued to diverge once overlapping route representations were no more similar than non-overlapping route representations? While speculative, it is possible that the divergence of hippocampal route representations may have been influenced by similarity among representations within input regions. For example, within PPA overlapping route representations tended to remain more similar than non-overlapping route representations even during the 2nd-half of learning.

While our findings are highly consistent with a differentiation account, it is important to emphasize that our results do not argue against a role for pattern separation, context learning or destination coding. Indeed, the lack of greater hippocampal pattern similarity among overlapping compared to non-overlapping routes at the beginning of learning may reflect an initial orthogonalization (pattern separation) of the route representations, with differentiation being additive to this initial orthogonalization. On the flip side, by establishing a role for hippocampal differentiation in a task closely modeled after rodent spatial learning paradigms, our results raise the important possibility that prior reports of pattern separation in the rodent hippocampus during spatial learning[1,2,19] might also reflect, at least in part, an influence of differentiation.

Beyond characterizing pattern similarity across voxels, a novel aspect of the present study is that we identified how responses within individual hippocampal voxels were reconfigured over the course of learning to produce pattern-level differentiation. Specifically, we observed a non-monotonic relationship in the similarly with which individual voxels responded to pairs of overlapping routes at the beginning vs. end of learning. Voxels that initially had 'moderate' levels of timecourse similarity were the voxels that exhibited the lowest timecourse similarity (or greatest differentiation) by the end of learning. This analysis was motivated by theoretical arguments and empirical evidence that plasticity in a competing memory's representation is non-monotonically related to its activation, with plasticity preferentially occurring when competing memories (or corresponding neural ensembles) are moderately activated[25,26,28]. While our analysis was inspired by this theoretical prediction, our analysis differed from prior examples in that we considered plasticity at the level of individual voxels as opposed to 'network level' memory strength or differentiation. Additionally, our analysis did not directly measure activation of competing memories; instead we used timecourse similarity as a way to infer when activation of competing neural ensembles was most likely to occur. Namely, we reasoned that timecourse similarity at the beginning of learning would be related to the degree of overlap among underlying neural ensembles (within a given voxel) and, therefore, to the probability that viewing one route would result in 'spill over' of activation to the competing neural ensemble. While our predictions were predicated on this assumption, the observed function was strongly consistent with the predicted non-monotonic relationship. To our knowledge, this is the first empirical evidence of non-monotonic plasticity within the hippocampus.

A final question concerns the nature of information coded for by the hippocampus. That is, what kinds of information were reflected in hippocampal activity patterns? Several prior studies have demonstrated spatial coding within the human hippocampus[44-48]. Here, our findings appear to be at odds with a spatial coding account in that hippocampal activity patterns diverged precisely when

routes were spatially overlapping. However, it is possible that when routes overlap the hippocampus forms separate spatial reference frames for each route. Thus, hippocampal activity patterns in the present study may have reflected spatial codes if overlapping routes were represented within distinct maps. Alternatively, the hippocampus may have coded for non-spatial information that differentiated the overlapping routes (e.g., pedestrians, cars, etc.). While we cannot determine the precise nature of hippocampal representations in the present study, it is clear that hippocampal representations were very distinct from representations in PPA–a cortical region adjacent to the hippocampus that has been implicated in coding spatial landmarks[49]. Indeed, PPA and the hippocampus exhibited opposite representational structures at the end of learning: in PPA, overlapping routes were more similar than non-overlapping routes, whereas in the hippocampus this representational structure was reversed. Thus, while the hippocampus may be part of a broader network of regions involved in spatial navigation and memory[9,35,45,48,50,51] our findings highlight unique representational dynamics within the hippocampus.

METHODS

Subjects

New York University (NYU) students and alumni who were familiar with the NYU campus participated in the study. Subjects were restricted to NYU alumni and students in order to facilitate route learning and to reduce potential between-subject variance. Subjects were between the ages of 18-35, right-handed, native English speakers, had normal or corrected-to-normal vision and had no history of neurological disorders. Twenty-two subjects participated in the behavioral experiment (15 female; mean age = 20.77). Two additional subjects' data were not collected due to technical errors. Twenty subjects (13 female; mean age = 22.15) participated in fMRI Experiment 1. Four additional subjects were excluded from data analysis - one for falling asleep in the scanner, two for technical errors during

scanning, and one due to unreliable localizer data (see Regions of Interest). Twenty-one subjects (9 female; mean age = 23.17) participated in fMRI Experiment 2. One additional subject's data was excluded from data analysis due to excessive head motion and another additional subject was excluded for technical errors with scanner. Sample sizes for the fMRI studies were based on a similar experiment from our lab[4]. Informed consent was obtained according to procedures approved by the New York University Committee on Activities Involving Human Subjects.

Stimuli and Design

In the behavioral experiment and fMRI Experiment 1 the stimuli consisted of eight routes that traversed the NYU campus (Fig. 1a). Each route was comprised of a series of 98 unique pictures. All pictures were taken at regular intervals (every 10 paces) from an egocentric perspective by a researcher walking along the route. All routes started in the same location and made exactly three turns before ending at distinct destinations. Critically, the 8 routes consisted of 4 overlapping pairs. Overlapping pairs followed the same path for the majority of the route before diverging on the third turn to their respective destinations. The pictures for each route were taken at different times and therefore the pictures during the overlapping portion of routes were subtly different and could be distinguished from one another based on subtle differences in the pedestrians, vehicles, lighting, etc. For analysis purposes, routes were divided into pairs that shared an overlapping path ('overlapping routes'; e.g. routes 1 and 2) or took distinct paths ('non-overlapping routes'; e.g. routes 1 and 3). Furthermore, each route was divided into two segments: 'Segment 1' refers to the segment of each route that overlapped with another route and 'Segment 2' refers to the route-unique segment of each route. The third turn–which marked the boundary between Segments 1 and 2–occurred at the exact same picture numbers within pairs or overlapping routes (e.g., for routes 1 and 2) and varied minimally (between picture numbers 74-77) across sets of overlapping pairs (e.g., for routes 1/2 vs. routes 3/4). Likewise, all turns within a pair of overlapping routes occurred at identical time points in order to maximize the similarity of overlapping routes. There was exactly one overlapping pair that left the

starting point in each cardinal direction (north, south, east, west). The 8 routes were divided into 2 sets (north/south routes and east/west routes). Each subject was assigned one set of routes (4 routes total) to learn, with the assignment of route sets alternating subject-by-subject. We included 2 sets of routes in order to ensure our results could not be explained by the idiosyncrasies of any one route.

A new set of 8 routes was used in fMRI Experiment 2 (Fig. 1b). The routes were constructed using the same parameters as the routes used in the behavioral and first fMRI experiments, with one key difference. Instead of all routes terminating at distinct locations, fMRI Experiment 2 contained pairs of routes that took distinct paths but ended at the same destination. As before, the 8 routes were divided into two sets of 4 and each set of 4 contained two pairs of overlapping routes. The routes in each set could be divided into pairs that (a) shared an overlapping path but terminated at distinct destinations ('overlapping routes'; e.g. routes 1 and 2), (b) had non-overlapping paths and terminated at distinct destinations ('non-overlapping routes'; e.g. routes 1 and 4) or (c) had non-overlapping paths but terminated at the same destinations ('same destination'; e.g. routes 1 and 3). Due to geographical constraints, the third turn (i.e. when overlapping routes diverged) in this set of routes occurred slightly later (between picture numbers 84-86) than in the set used in the behavioral and first fMRI experiments.

Videos of the overlapping route pairs used in the experiments are available in Supplementary Videos 1-8.

Procedure

Behavioral Experiment

Route Learning: Subjects completed 14 rounds of route learning, with each route presented twice per round in random order. During a route learning trial, pictures from a route were presented in rapid succession (220 ms per picture, 10 ms blank screen in between pictures). Importantly, subjects were not told the destination of

the route prior to the trial. Rather, the destination was only revealed at the end of the route, with the final picture (the destination) presented for 1690 ms. The destination's name was also displayed above the final picture. Each route learning trial lasted a total of 24s and was followed by a 1-s inter-trial interval (ITI) during which a fixation cross was presented. Each round also contained two 'catch' trials to ensure subjects' vigilance but were excluded from all analyses. For each catch trial, a route began as with a normal trial but the presentation stopped at a pre-selected picture number. A cue then appeared above the picture either instructing participants to identify (1) the routes' final destination (destination test) or (2) the direction of the next turn (direction test). During the 3s response period the picture and test cue remained on screen with the four destination labels (destination test) or left/right labels (direction test) printed below the picture and participants selected their response using a keyboard. Catch trials stopped on pictures presented between 3-15s after the trial onset and at intervals of 1.5 s (to coincide with the TR length in the fMRI experiments; see fMRI Acquisition). The combined duration of the two test trials within each round were constrained to equal the duration of a full route learning trial (24 seconds). Although each subject completed an equal number of destination and direction catch trials throughout the experiment, and each route was tested an equal number of times, the assignment of catch trial type to both route number and round was randomized so as not to be predictable. That is, within a given round there could be 2 destination catch trials, 2 direction catch trials, or 1 of each, and a given route could be tested twice via a destination catch trial, twice as a direction catch trial, or once as each test.

Inter-Round Picture Test: At the end of each of the 14 route learning rounds, subjects were shown 20 static pictures, one at a time, drawn from the routes (5 per route in random order) and for each picture subjects were asked to select the corresponding destination from a set of four label options. The inter-round picture test was self-paced and subjects responded via keyboard. To ensure that the five pictures tested from each route in each test round were evenly sampled across positions in the route, each route's 95 pictures (excluding the last 3 pictures that

contained visuals of the destination) were divided into 5 time-bins of 19 pictures. For each inter-round picture test, one picture from each time-bin, from each route, was randomly selected to be tested with the constraint that a given picture was only tested once throughout the experiment. Responses on the test were divided into three groups: (1) 'target' if subjects selected the correct destination, (2) 'competitor' if subjects selected the overlapping route's destination, and (3) 'other' if subjects selected the destination from a non-overlapping route.

Map Test: In order to assess each subject's knowledge of the routes, subjects also completed a map test after finishing all rounds of route learning. For each trial on the map test, subjects were cued with a picture of a route's destination for 4s. A map of the NYU campus then appeared on screen and subjects had 8s to click on the spatial location of the cued destination using a computer mouse. They were then prompted to draw with a pen the route taken to that destination on a paper print out of the campus map. Finally, participants completed both the Santa Barbara Sense of Direction Scale (SBSOD) and the Questionnaire on Spatial Representation (QSR) to assess their spatial acuity and reasoning. Results from the map test and questionnaires are not reported in the current study.

fMRI Experiments 1 and 2

Route Learning: The procedures from the behavioral experiment were slightly modified to be suitable for fMRI scanning. In both fMRI experiments, subjects first completed 2 practice route learning rounds (2 repetitions of each route per round) to familiarize them with the routes and task structure. Subjects then entered the scanner and completed an additional 14 rounds of route learning. The practice rounds were identical to the scanner rounds except that the first practice round did not contain any catch trials. During the scanned route learning rounds, the ITI was 6 s (fixation cross) to allow for better separation of hemodynamic response functions.

Inter-Round Picture Test: The inter-round picture test used in the fMRI experiments was shorter than in the behavioral experiment. In the fMRI version, there were a total of only 4 trials which contained pictures randomly sampled from

the 4 routes. The sampled pictures were not constrained to be from different routes. The only constraint was that the pictures used in the inter-round picture test were not used in the post-scan memory test (described below). Additionally, in the fMRI version of the inter-round picture test subjects were shown each picture for a fixed amount of time (2.5 s) and could only respond during that time, using an MRI-compatible button box. Because the inter-round picture tests in the fMRI experiments only sparsely assessed route learning, these data are not reported. These test trials were only included to motivate subjects to learn the routes.

Functional Localizer: Following the 14 rounds of route learning subjects completed one localizer scan that was used to functionally define regions of interest for the fMRI analyses. The localizer scan contained 36 alternating blocks of three image types (12 blocks per category): faces, scenes (hallways or houses), and objects (cars or guitars). Each block lasted a total of 6 s and contained 12 greyscale images presented for 500 ms each. Subjects pressed a button whenever they detected a scrambled image, which occurred on half of all blocks (counterbalanced across category). An additional 12 baseline 'blocks' showing a blank grey screen (also 6 s each) were randomly interspersed with the other blocks.

Post Tests: After exiting the scanner subjects first completed a map test (identical to the behavioral experiment). Next, subjects completed an extended picture test which included ten pictures drawn from each route (every 10th picture from picture 4 to 94), tested in random order. On each trial, the route picture was presented above the set of destination names (4 destination names in Experiment 1 and 2 destination names in Experiment 2). Subjects used a computer mouse to click on the destination name associated with each picture. This test was self-paced. Finally, subjects completed the Santa Barbara Sense of Direction Scale (SBSOD) and the Questionnaire on Spatial Representation (QSR).

fMRI Data Analysis

MRI Acquisition: Scanning was performed on a 3T Siemens Allegra head-only scanner at the Center for Brain Imaging at New York University using a Siemens

head coil. Structural images were collected using a T1-weighted protocol (256 × 256 matrix, 176 1-mm sagittal slices). Functional images were acquired using a T2* weighted EPI single shot sequence containing 26 contiguous axial slices oriented parallel to the long-axis of the hippocampus (repetition time = 1.5 s, echo time = 23 ms, flip angle = 77 degrees, voxel size = 2 x 2 x 2 mm). The functional images did not cover the entire brain; rather, a limited field of view centered on the hippocampus was chosen in order to improve spatial resolution of data from the hippocampus. For the route learning scans, the first 6 volumes (during which time a "Get Ready" screen was presented, followed by a fixation cross) were discarded to account for T1 stabilization. For the localizer scan, the first 8 volumes and last 8 volumes (during which time a fixation cross was presented) were discarded. Field map and calibration scans were collected to improve functional-to-anatomical coregistration.

fMRI Preprocessing: Images were preprocessed using SPM8 (Wellcome Department of Cognitive Neurology, London, United Kingdom), FSL (FMRIB's Software Library, Oxford,United Kingdom) and custom Matlab (The MathWorks, Natick, MA) routines. The preprocessing procedures included correction for head motion, coregistration of functional to anatomical images (using a registration procedure that aligned both functional and anatomical images to a calibration scan), and an unwarping procedure. Images from the functional localizer scan were spatially smoothed using a 4-mm full-width/half-maximum Gaussian kernel. Images from the route learning phase, which were used for pattern analyses, were smoothed using a moderate 2-mm full-width/half-maximum Gaussian kernel in order to improve signal-to-noise ratio. Prior research suggests that smoothing does not reduce sensitivity of pattern-based fMRI analyses [50]. All analyses were performed in subjects' native space.

fMRI univariate analysis: To analyse the localizer data, SPM was used to construct a general linear model with three regressors of interest corresponding to the three visual categories (scenes, faces, objects). These regressors were constructed as boxcar functions that onset at the first image of a category block and lasted for the duration of the block. Motion, block, and linear drift were modeled as regressors of

no interest. All regressors were convolved with a canonical double-gamma hemodynamic response function. A linear contrast of scenes vs. faces and objects was used to obtain voxelwise estimates of scene sensitivity and a linear contrast of faces, scenes, and objects vs. baseline was used to obtain voxelwise estimates of visual sensitivity.

Regions of interest: Anlyses were performed using a region of interest (ROI) approach targeting the hippocampus, parahippocampal place area (PPA), and retrosplenial cortex (RSC). Anatomical hippocompal ROIs were defined using freesurfer's automated sub-cortical segmentation procedure. The resultant hippocampal ROIs were then visually inspected and manually edited for any inaccuracies before registering them to each subject's functional space. In order to identify voxels with high signal-to-noise ratios and to create ROI masks the same size as the PPA and RSC masks (see below), the hippocampal ROI consistend of the top 300 visually-responsive voxels within bilateral hippocampus, as determined from the category localizer (contrast of faces, scenes, and objects vs. baseline). Although this voxel selection procedure was implemented to increase our sensitivity to detect small differences in hippocampal patterns, it is important to note that our main findings were not dependent on such selection methods. Indeed when no voxel selection was applied within the hippocampus the interaction between overlap (overlap/non-overlap) and learning (1st half/2nd half) remained significant ($F(1,39) = 4.75$, $P = 0.0354$), as did the reversal effect in the 2nd half of learning ($F(1,39) = 7.30$, $P = 0.0102$).

PPA and RSC were identified using a combination of the category localizer and group-based probabilistic scene-selective ROIs identified from previous studies [51]; http://web.mit.edu/bcs/nklab/GSS.shtml). First, the group-based PPA and RSC masks were registered to each subject's native space and voxels overlapping with the anatomically defined hippocampal masks were removed from the PPA/RSC masks to ensure independent ROIs. Then, the top 300 scene-selective voxels (contrast of scenes vs. faces and objects from the category localizer) within PPA and, separately, within RSC, were selected. This method ensured that the PPA and RSC

ROIs were subject-specific but equal in size (number of voxels) and general location across all subjects [47]. Note: we chose 300 voxels as an a priori threshold for all our ROIs. This number corresponded to roughly the top 20% of the hippocampal voxels, 30% of the voxels within the group-based PPA mask, and 15% of the voxels in the group-based RSC mask. One subject from Experiment 1 was excluded because the average *t* value within their PPA ROI was more than two standard deviations below the mean PPA response in Experiment 1 (this was the only subject with a mean PPA or RSC response that was more than 2 standard deviations below the experiment mean); subjective assessment of data from this subject confirmed that there was no well-defined cluster within the group-based PPA mask that selectively responded to scenes.

Spatiotemporal pattern similarity: Pattern similarity analyses were performed on 'raw' (unmodeled) fMRI data. Several additional preprocessing steps were performed prior to performing pattern analyses. Functional images were detrended, high-pass filtered (0.01 Hz), and then z-scored within run. For route learning trials, volumes 3-19 (corresponding to 3-27s after stimulus onset) were divided into volumes corresponding to Segment 1 (i.e. the portion of each route that shared a path with another route) and Segment 2 (i.e. the unique portion of each route after overlapping paths diverged). The volume in each route corresponding to the transition between Segments 1 and 2 (i.e., the third turn in the routes) was discarded from analyses in order to keep Segments 1 and 2 distinct. In Experiment 1, Segment 1 occurred within the first 11 volumes and Segment 2 occurred within the last 4 volumes. In Experiment 2 the overlapping routes diverged slightly later; thus, Segment 1 corresponded to the first 12 volumes Segment 2 corresponded to the last 3 volumes. To perform pattern analyses, spatial activity patterns were concatenated across volumes of interest so that each route Segment was represented by a spatiotemporal pattern of activity whose vector length was equal to the number of voxels within an ROI * the number of TRs included in the Segment.

For each subject and each ROI, we computed pattern similarity scores (Pearson correlations) reflecting the representational similarity across each pair of

routes. Correlations were always performed using data from independent fMRI runs (odd and even runs) in order to ensure independence. Thus, for analysis of data from the first half of learning, each route's average spatiotemporal activity pattern was obtained from runs 1, 3, and 5 (odd runs) and, separately, from runs 2, 4, and 6 (even runs); average 'odd run patterns' were then correlated with average 'even run patterns.' Likewise, for analysis of data from the second half of learning, each route's average spatiotemporal activity pattern was obtained from runs 9, 11, and 13 (odd runs) and, separately, from runs 10, 12, and 14 (even runs), and odd and even patterns were correlated. Data from runs 7 and 8 were excluded in order to ensure an equal number of odd and even runs within each half. Because each subject studied 4 routes, a 4 x 4 correlation matrix was generated for each subject (Fig. 3a). Before any correlation values were averaged within conditions (e.g., overlapping routes), correlation coefficients were z-transformed (Fisher's z).

Timecourse similarity: Timecourse similarity indexed the degree to which individual voxels were 'shared' across a given pair of routes. To compute timecourse similarity, we first obtained route-specific vectors of activation (using Segment 1 data only) for each voxel. The length of each timecourse vector was equal to the number of Segment 1 TRs (11 in Experiment 1; 12 in Experiment 2). Timecourse vectors were separately averaged across odd and even runs within each half (as with the spatiotemporal pattern analyses). Average timecourse vectors were then correlated (Pearson correlation) for every pair of routes, separately for each learning half (Fig. 4b). Resulting correlation coefficients were z-transformed (Fisher's z).

Curve Fitting Procedure: To estimate the relationship between timecourse similarity for overlapping routes at the beginning vs. end of learning, we used the probabilistic curve induction and testing (P-CIT) Bayesian curve-fitting toolbox [34];http://code.google.com/p/p-cit-toolbox). To prepare the data for the algorithm, the 300 voxels within each subject's hippocampal ROI were rank-ordered according to 1st half timecourse similarity values and then divided into 60 bins (5 voxels/bin) ordered from lowest to highest similarity. Second half timecourse similarity values

were then obtained for each of these 60 bins. This process was independently repeated for each pair of overlapping routes and the resulting data (i.e., the sets of 60 2nd half timecourse similarity values) were then averaged across route pairs, producing a single set of 60 values per subject. Consistent with all prior implementations of the P-CIT toolbox [28,33,34], we then aggregated data across subjects (i.e., 41 subjects x 60 voxel bins = 2,460 data points) and the P-CIT algorithm was applied to these aggregated data.

The P-CIT algorithm approximates the posterior distribution over possible plasticity curves (i.e which curves are most probable given the data). As a first step, the algorithm rescales the predictor variable (here, 1st half timecourse similarity values) to fit within a range of 0 to 1. Next, the algorithm defines a parameterized family of curves (piecewise-linear curves with three segments) and then randomly samples 50,000 curves from this parameterized space in which the bounds of the x-axis extend from 0 to 1 and the bounds on the y-axis extend from -1 to 1. For each sampled curve the algorithm assigns an importance weight indicating how well the curve's shape fit the observed data. After assigning importance weights to all 50,000 samples, the algorithm regenerates a new set of sample curves by taking the highest weighted curves from the previous sample and slightly distorting them. This process of assigning importance weights to sampled curves and resampling curves based on those importance weights was repeated for 20 iterations. The resultant collection of weighted samples can be interpreted as an approximate posterior probability distribution over curves. In other words, curve samples with high importance weights are relatively probable given the data whereas sample curves with low importance weights are relatively improbable. The algorithm then estimates a curve by averaging the sampled curves together weighted by their importance values. Credible intervals were computed by finding the range of y values contained within the middle 90% of the curve probability mass at each interval along the x-axis. To summarize the level of evidence in favor of the non-monotonic plasticity hypothesis the algorithm computes the log Bayes factor that reflects the log of the ratio of evidence in favor of the hypothesis to the evidence

against the hypothesis. To compute the log Bayes factor, first each curve in the final set of samples was labeled as either theory consistent (i.e. if moving from left to right the curve dropped below its starting point and then rose above its starting value) or theory inconsistent. Then, the proportion of the posterior probability distribution mass taken up by theory-consistent samples was computed by summing together the weights corresponding to theory consistent curves and dividing that value by the sum of the weights corresponding to theory-inconsistent curves[28]. The log Bayes factor was then computed as:

$$\boldsymbol{f} = \mathbf{ln}\left(\frac{\boldsymbol{sum\ of\ importance\ weights\ for\ theory\ consistent\ curves}}{\boldsymbol{sum\ of\ importance\ weights\ for\ theory\ inconsistent\ curves}} \div \boldsymbol{correction\ factor}\right)$$

where

$$\boldsymbol{correction\ factor} = \frac{\%\ \boldsymbol{of\ the\ space\ of\ possible\ curves\ that\ is\ theory\ consistent}}{\%\ \boldsymbol{of\ the\ space\ of\ possible\ curves\ that\ is\ theory\ inconsistent}} = \frac{\mathbf{.3328}}{\mathbf{.6672}} = \mathbf{0.4988}$$

The correction factor compensates for the fact that only 33.28% of the space of possible curves is comprised of theory consistent curves. Without this adjustment we would expect a large imbalance in favor theory-inconsistency due to chance. To test whether the obtained log Bayes factor value could have arisen due to chance, we ran a non-parametric permutation test. Specifically, we re-ran the algorithm 500 times with the 1st and 2nd half timecourse similarity values shuffled within subject yielding an empirical null distribution over log Bayes factor values. We then measured where our observed log Bayes factor value fell on this distribution to compute the probability of obtaining this value under the null hypothesis. Lastly, the P-CIT toolbox also calculates a chi-squared likelihood ratio test that assesses how well the estimated curve explains the data. The *P*-value for this test indicates the probability of obtaining the observed predictive accuracy relative to a null model in which there is no relationship between timecourse similarity values across halves.

Statistics: For all behavioral and fMRI analyses (with the exception of the curve-fitting procedures) we used standard random-effects statistics (paired sample *t*-tests and repeated measures ANOVA). Two-tailed tests were used throughout at an

alpha threshold of 0.05. Unless otherwise noted, analyses combined data across experiments 1 and 2. For all ANOVAs run on these combined data, experiment number was included as a factor. For all of the hippocampal ANOVA effects described in the main text, interactions with experiment number were not significant (*P*s > 0.2). See Supplementary Figs 2 and 6 for hippocampal and PPA data separated by experiment. All statistical models were implemented using R (https://www.r-project.org).

Statistics: For all behavioral and fMRI analyses (with the exception of the curve-fitting procedures) we used standard random-effects statistics (paired sample *t*-tests and repeated measures ANOVA). Two-tailed tests were used throughout at an alpha threshold of 0.05. Unless otherwise noted, analyses combined data across Experiments 1 and 2. For all ANOVAs run on these combined data, experiment number was included as a factor. For all of the hippocampal ANOVA effects described in the main text, interactions with experiment number were not significant (*P*s > 0.2). See Supplementary Figs 2 and 6 for hippocampal and PPA data separated by experiment. All statistical models were implemented using R (https://www.r-project.org).

Data and Code Availability: All relevant data and code are available upon request from the first author (avi.chanales@nyu.edu).

Acknowledgements: We thank Anthony Stigliani and Kalanit Grill-Spector for providing stimuli for the category localizer. This work was supported by a grant from the National Institutes of Health (1RO1NS089729) to B.A.K.

REFERENCES

1. Wood, E. R., Dudchenko, P. A., Robitsek, R. & Eichenbaum, H. Hippocampal neurons encode information about different types of memory episodes occurring in the same location. *Neuron* **27,** 623–633 (2000).

2. Frank, L. M., Brown, E. N. & Wilson, M. Trajectory encoding in the hippocampus and entorhinal cortex. *Neuron* **27,** 169–178 (2000).

3. Hsieh, L.-T., Gruber, M. J., Jenkins, L. J. & Ranganath, C. Hippocampal activity patterns carry information about objects in temporal context. *Neuron* **81,** 1165–1178 (2014).

4. Favila, S. E., Chanales, A. J. H. & Kuhl, B. A. Experience-dependent hippocampal pattern differentiation prevents interference during subsequent learning. *Nat. Commun.* **7** (2016).

5. Chadwick, M. J., Hassabis, D. & Maguire, E. A. Decoding overlapping memories in the medial temporal lobes using high-resolution fMRI. *Learn. Mem.* **18,** 742–746 (2011).

6. Ginther, M. R., Walsh, D. F. & Ramus, S. J. Hippocampal neurons encode different episodes in an overlapping sequence of odors task. *J. Neurosci.* **31,** 2706–2711 (2011).

7. Agster, K. L., Fortin, N. J. & Eichenbaum, H. The hippocampus and disambiguation of overlapping sequences. *J. Neurosci.* **22,** 5760–5768 (2002).

8. McKenzie, S. *et al.* Hippocampal representation of related and opposing memories develop within distinct hierarchically organized neural schemas. *Neuron* **83,** 202–215 (2014).

9. Brown, T. I., Ross, R. S., Keller, J. B., Hasselmo, M. E. & Stern, C. E. Which way was i going? contextual retrieval supports the disambiguation of well learned overlapping navigational routes. *J. Neurosci.* **30,** 7414–7422 (2010).

10. Kumaran, D. & Maguire, E. A. The dynamics of hippocampal activation during encoding of overlapping sequences. *Neuron* **49,** 617–629 (2006).

11. Gilbert, P. E., Kesner, R. P. & Lee, I. Dissociating hippocampal subregions: A double dissociation between dentate gyrus and CA1. *Hippocampus* **11,** 626–636 (2001).

12. Schlichting, M. L., Mumford, J. A. & Preston, A. R. Learning-related representational changes reveal dissociable integration and separation signatures in the hippocampus and prefrontal cortex. *Nat. Commun.* **6,** 8151 (2015).

13. Grieves, R. M., Wood, E. R. & Dudchenko, P. A. Place cells on a maze encode routes rather than destinations. *eLife* **5,** (2016).

14. Kyle, C. T., Stokes, J. D., Lieberman, J. S., Hassan, A. S. & Ekstrom, A. D. Successful retrieval of competing spatial environments in humans involves hippocampal pattern separation mechanisms. *eLife* **4,** (2015).

15. OReilly, R. C. & McClelland, J. L. Hippocampal conjunctive encoding storage, and recall: Avoiding a trade-off. *Hippocampus* **4,** 661–682 (1994).

16. OReilly, R. C. & Rudy, J. W. Conjunctive representations in learning and memory: Principles of cortical and hippocampal function. *Psychol. Rev.* **108,** 311–345 (2001).

17. Marr, D. Simple memory: a theory for archicortex. *Phil. Trans. R. Soc. B* **262,** 23–81 (1971).

18. Leutgeb, S., Leutgeb, J. K., Treves, A., Moser, M.-B. & Moser, E. I. Distinct ensemble codes in hippocampal areas ca3 and ca1. *Science* **305,** 1295–1298 (2004).

19. Leutgeb, J. K., Leutgeb, S., Moser, M.-B. & Moser, E. I. Pattern separation in the dentate gyrus and ca3 of the hippocampus. *Science* **315,** 961–966 (2007).

20. McHugh, T. J. *et al.* Dentate gyrus nmda receptors mediate rapid pattern separation in the hippocampal network. *Science* **317,** 94–99 (2007).

21. Bakker, A., Kirwan, C. B., Miller, M. & Stark, C. E. L. Pattern separation in the human hippocampal CA3 and dentate gyrus. *Science* **319,** 1640–1642 (2008).

22. Yassa, M. A. & Stark, C. E. Pattern separation in the hippocampus. *Trends Neurosci.* **34,** 515–525 (2011).

23. Treves, A. & Rolls, E. T. Computational analysis of the role of the hippocampus in memory. *Hippocampus* **4,** 374–391 (1994).

24. Norman, K. A., Newman, E., Detre, G. & Polyn, S. How inhibitory oscillations can train neural networks and punish competitors. *Neural Comput.* **18,** 1577–1610 (2006).

25. Norman, K. A., Newman, E. L. & Detre, G. A neural network model of retrieval-induced forgetting. *Psychol. Rev.* **114,** 887–953 (2007).

26. Hulbert, J. C. & Norman, K. A. Neural differentiation tracks improved recall of competing memories following interleaved study and retrieval practice. *Cereb. Cortex* **25,** 3994–4008 (2014).

27. Poppenk, J. & Norman, K. A. Briefly cuing memories leads to suppression of their neural representations. *J. Neurosci.* **34,** 8010–8020 (2014).

28. Lewis-Peacock, J. A. & Norman, K. A. Competition between items in working memory leads to forgetting. *Nat. Commun.* **5,** 5768 (2014).

29. Schapiro, A. C., Kustner, L. V. & Turk-Browne, N. B. Shaping of object representations in the human medial temporal lobe based on temporal regularities. *Curr. Biol.* **22,** 1622–1627 (2012).

30. Norman, K. A. & OReilly, R. C. Modeling hippocampal and neocortical contributions to recognition memory: a complementary-learning-systems approach. *Psychol. Rev.* **110,** 611–646 (2003).

31. Epstein, R., Harris, A., Stanley, D. & Kanwisher, N. The parahippocampal place area. *Neuron* **23,** 115–125 (1999).

32. Epstein, R. A. Parahippocampal and retrosplenial contributions to human spatial navigation. *Trends Cogn. Sci.* **12,** 388–396 (2008).

33. Lee, I., Griffin, A. L., Zilli, E. A., Eichenbaum, H. & Hasselmo, M. E. Gradual translocation of spatial correlates of neuronal firing in the hippocampus toward prospective reward locations. *Neuron* **51,** 639–650 (2006).

34. Wikenheiser, A. M. & Redish, A. D. Hippocampal theta sequences reflect current goals. *Nat. Neurosci.* **18,** 289–294 (2015).

35. Brown, T. I. *et al.* Prospective representation of navigational goals in the human hippocampus. *Science* **352,** 1323–1326 (2016).

36. Davachi, L. & DuBrow, S. How the hippocampus preserves order: the role of prediction and context. *Trends Cogn. Sci.* **19,** 92–99 (2015).

37. Olafsdottir, H. F., Barry, C., Saleem, A. B., Hassabis, D. & Spiers, H. J. Hippocampal place cells construct reward related sequences through unexplored space. *eLife* **4,** (2015).

38. Hindy, N. C., Ng, F. Y. & Turk-Browne, N. B. Linking pattern completion in the hippocampus to predictive coding in visual cortex. *Nat. Neurosci.* **19,** 665–667 (2016).

39. Wimber, M., Alink, A., Charest, I., Kriegeskorte, N. & Anderson, M. C. Retrieval induces adaptive forgetting of competing memories via cortical pattern suppression. *Nat. Neurosci.* **18,** 582–589 (2015).

40. Newman, E. L. & Norman, K. A. Moderate excitation leads to weakening of perceptual representations. *Cereb. Cortex* **20,** 2760–2770 (2010).

41. Kim, G., Lewis-Peacock, J. A., Norman, K. A. & Turk-Browne, N. B. Pruning of memories by context-based prediction error. *Proc. Natl. Acad. Sci. USA* **111,** 8997–9002 (2014).

42. Detre, G. J., Natarajan, A., Gershman, S. J. & Norman, K. A. Moderate levels of activation lead to forgetting in the think/no-think paradigm. *Neuropsychologia* **51,** 2371–2388 (2013).

43. LaRocque, K. F. *et al.* Global similarity and pattern separation in the human medial temporal lobe predict subsequent memory. *J. Neurosci.* **33,** 5466–5474 (2013).

44. Miller, J. F. *et al.* Neural activity in human hippocampal formation reveals the spatial context of retrieved memories. *Science* **342,** 1111–1114 (2013).

45. Doeller, C. F., King, J. A. & Burgess, N. Parallel striatal and hippocampal systems for landmarks and boundaries in spatial memory. *Proc. Natl. Acad. Sci. USA* **105,** 5915–5920 (2008).

46. Nielson, D. M., Smith, T. A., Sreekumar, V., Dennis, S. & Sederberg, P. B. Human hippocampus represents space and time during retrieval of real-world memories. *Proc. Natl. Acad. Sci. USA* **112,** 11078–11083 (2015).

47. Hassabis, D. *et al.* Decoding neuronal ensembles in the human hippocampus. *Curr. Biol.* **19,** 546–554 (2009).

48. Ekstrom, A. D. *et al.* Cellular networks underlying human spatial navigation. *Nature* **425,** 184–188 (2003).

49. Marchette, S. A., Vass, L. K., Ryan, J. & Epstein, R. A. Outside looking in: landmark generalization in the human navigational system. *J. Neurosci.* **35,** 14896–14908 (2015).

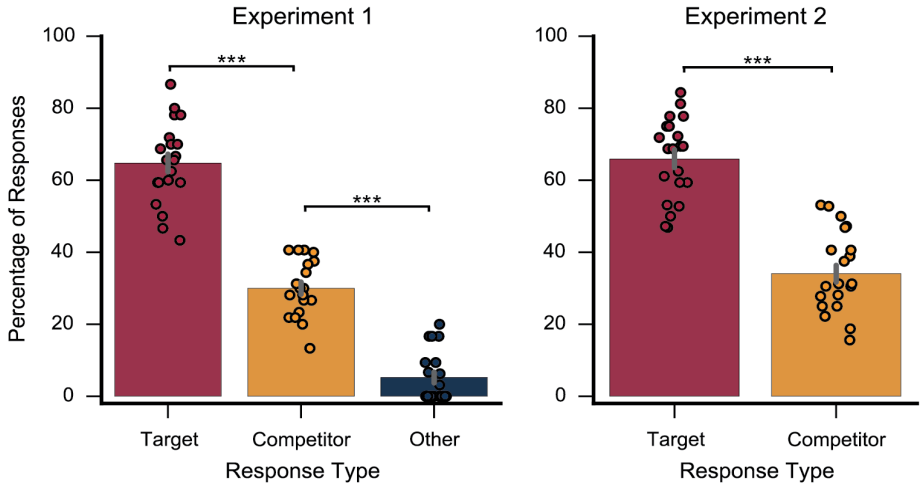
50. Spiers, H. J. & Barry, C. Neural systems supporting navigation. *Curr. Opin. Behav. Sci.* **1,** 47–55 (2015).

51. Burgess, N., Maguire, E. A. & OKeefe, J. The human hippocampus and spatial and episodic memory. *Neuron* **35,** 625–641 (2002).

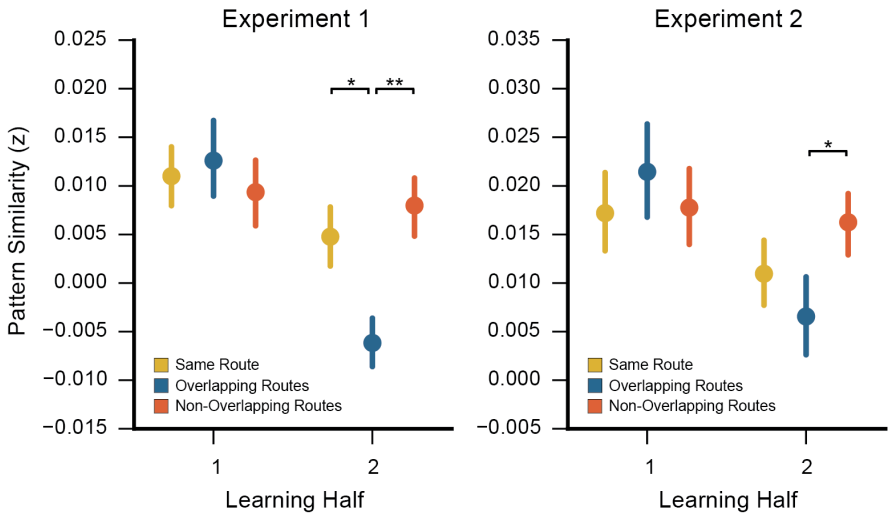
52. Beeck, H. P. O. de. Against hyperacuity in brain reading: spatial smoothing does not hurt multivariate fMRI analyses? *NeuroImage* **49,** 1943–1948 (2010).

53. Julian, J., Fedorenko, E., Webster, J. & Kanwisher, N. An algorithmic method for functionally defining regions of interest in the ventral visual pathway. *NeuroImage* **60,** 2357–2364 (2012).

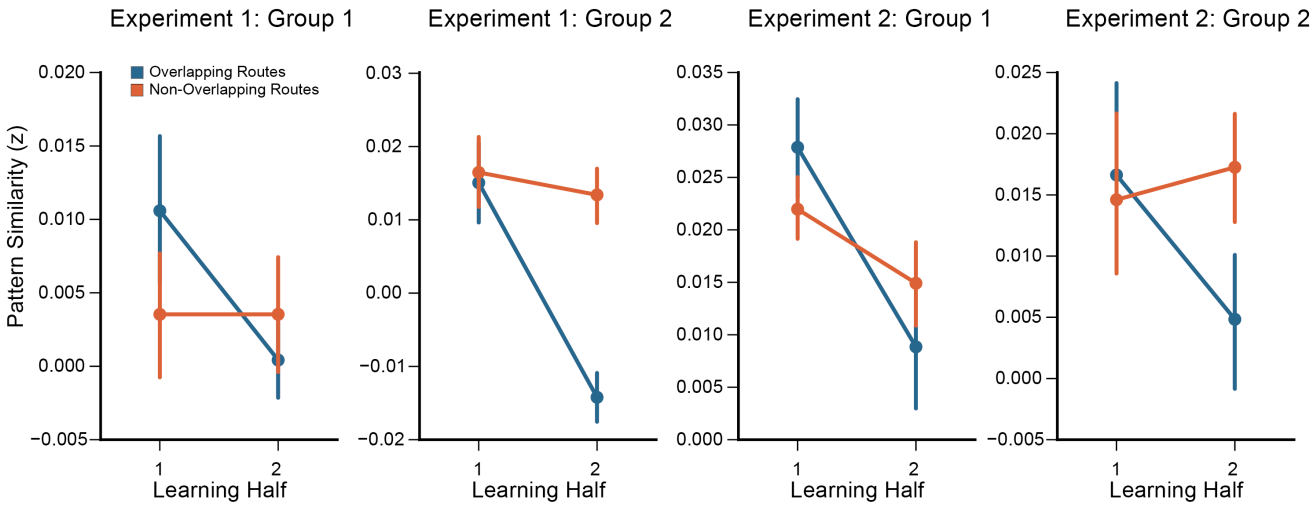
Supplement



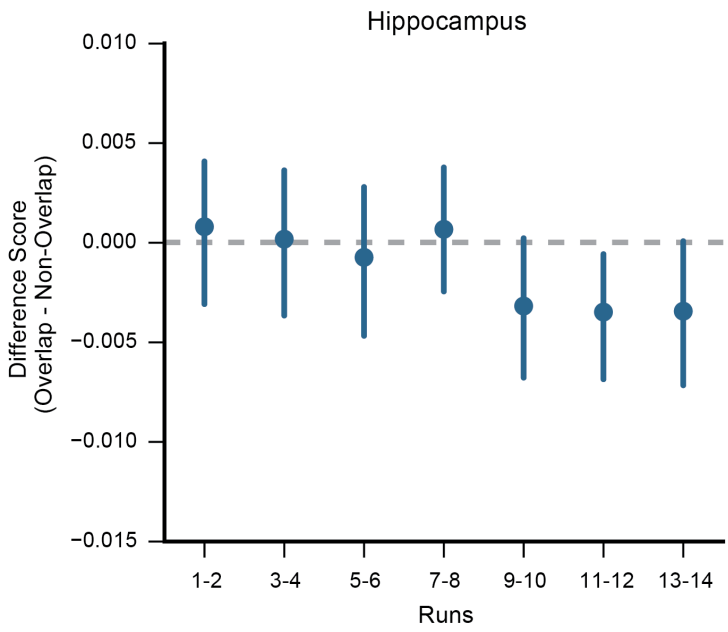
Supplementary Figure 1. Behavioral results from the post-scan picture test in each fMRI experiment. After finishing all 14 learning rounds and the localizer scan subjects exited the scanner and completed the picture test. On each trial, subjects were shown a static picture drawn from the route stimuli (see Methods). Directly below each picture was a set of destination names. Subjects were instructed to select the destination name corresponding to the route picture. In Experiment 1, each trial had 4 destination options corresponding to the target destination, the overlapping route destination ('competitor') and the two non-overlapping route destinations ('other'). In Experiment 2 all of the routes studied by each subject ended in one of two possible destinations. Therefore, on each trial the two destination options corresponded to either the target destination or the overlapping route destination ('competitor'). Analyses were restricted to pictures drawn from Segment 1 of each route (i.e., the segments that contained overlap) in order to test discrimination of the overlapping routes. In both experiments subjects successfully learned to discriminate between the overlapping routes as evidenced by a higher percentage of target responses than competitor responses (Experiment 1: $t_{19} = 8.59$, $P = 0.00000006$; Experiment 2: $t_{20} = 6.44$, $P = 0.000003$). In Experiment 1, subjects were more likely to select the competitor destination than one of the 'other' destinations ($t_{19} = 11.52$, $P < 0.00000001$), indicating that route overlap contributed to memory interference. Error bars reflect +/- SEM. *** $P < 0.001$



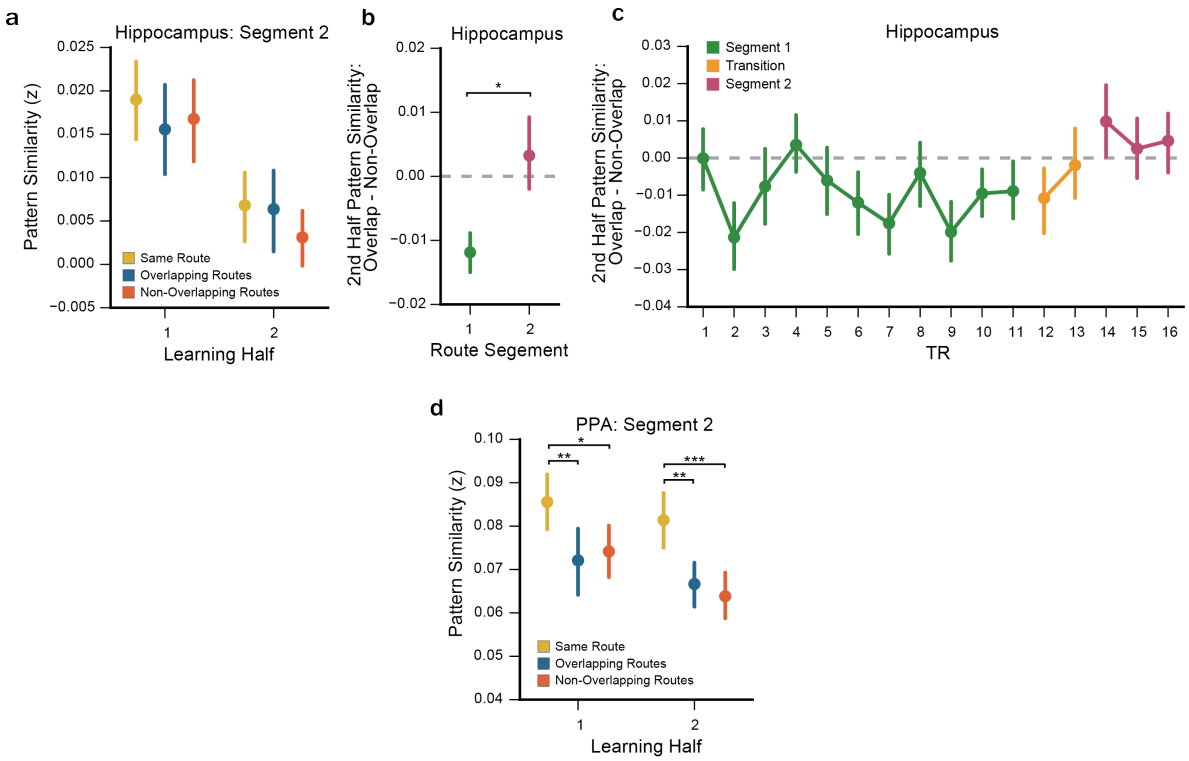
Supplementary Figure 2. Learning-related changes in hippocampal pattern similarity (Segment 1 only) for each fMRI experiment. In each fMRI experiment, overlapping route similarity decreased across learning halves, relative to non-overlapping route similarity (Experiment 1: $F_{1,19} = 5.99$, $P = 0.024$; Experiment 2: $F_{1,20} = 8.02$, $P = 0.010$). Furthermore, the reversal effect (overlapping route similarity < non-overlapping route similarity) was also significant in the 2nd half of learning for both experiments (Experiment 1: $t_{19} = 3.03$, $P = 0.007$; Experiment 2: $t_{20} = 2.28$, $P = 0.034$). Error bars reflect +/- SEM. * $P < 0.05$, ** $P < 0.01$



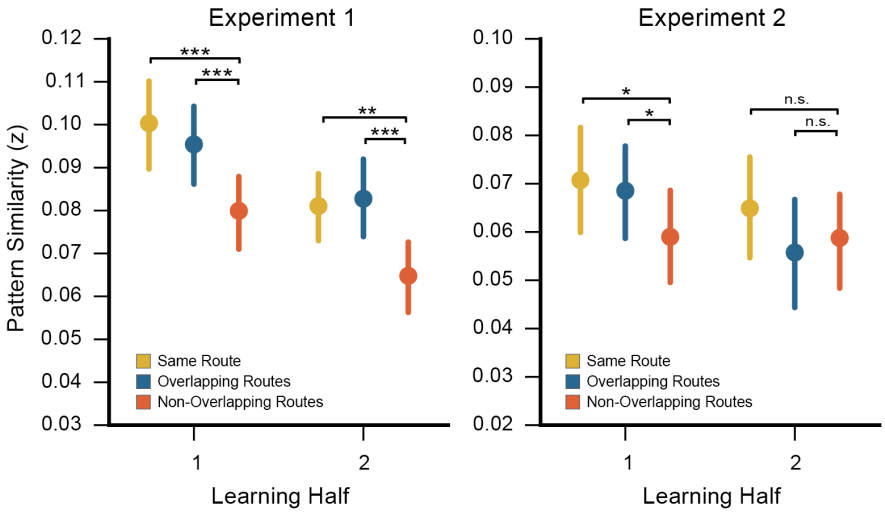
Supplementary Figure 3. Learning-related changes in hippocampal spatiotemporal pattern similarity (Segment 1 only) for each stimulus group. Across the two experiments, 4 different groups of route stimuli were used: 2 unique sets per experiment. Within each experiment, the two sets were divided between odd and even subject numbers. Within all 4 stimuli groups, the qualitative pattern was the same: similarity among the overlapping routes decreased relative to similarity among non-overlapping routes. For each stimulus group, in the second half of learning overlapping route similarity was numerically lower than non-overlapping route similarity (reversal effect).



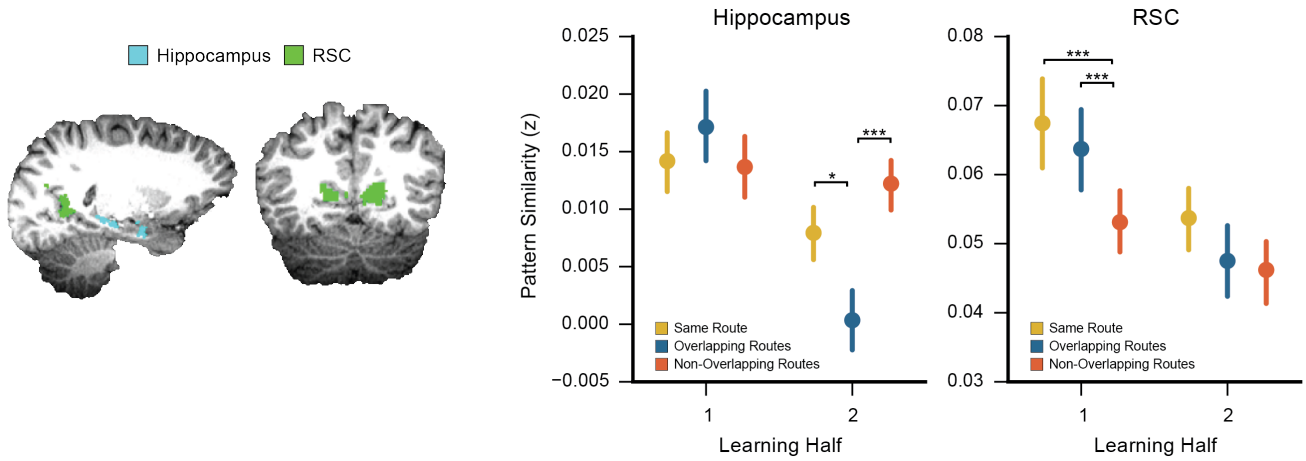
Supplementary Figure 4. Hippocampal spatiotemporal pattern similarity (Segment 1 only) computed every two runs. Qualitatively, there was no evidence for a reversal effect (overlapping route similarity < non-overlapping route similarity) until run 9. Note: each run contained two repetitions of each route.



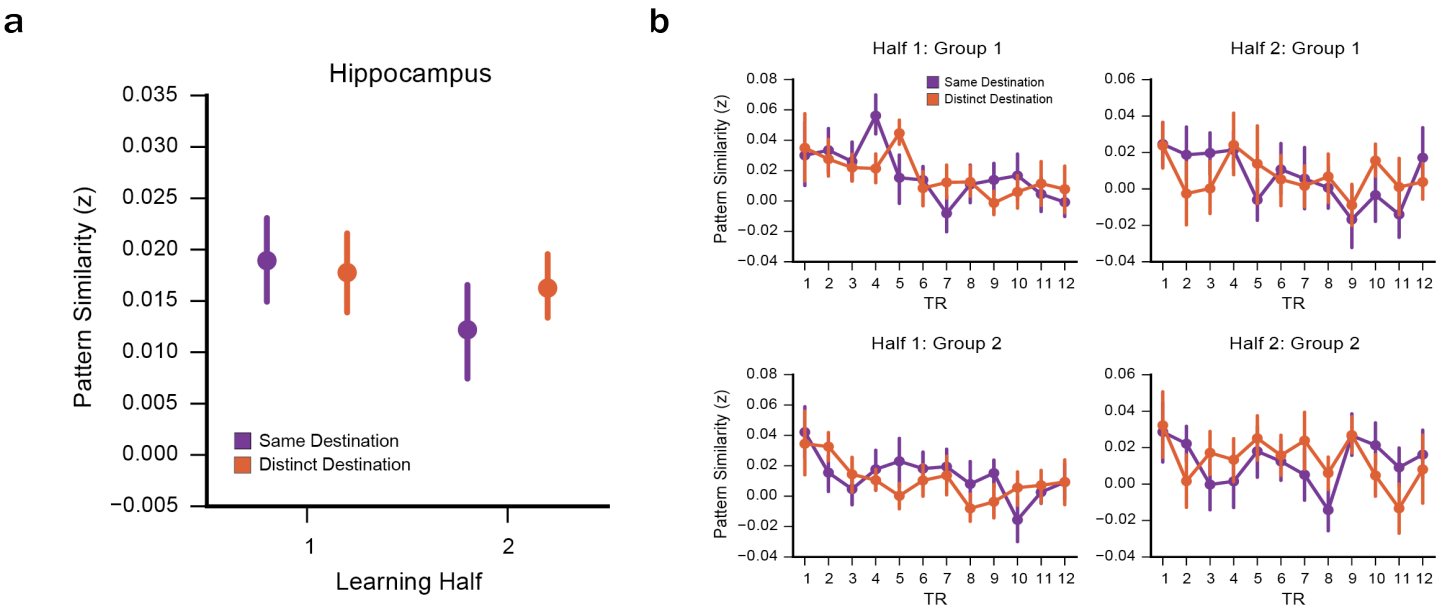
Supplementary Figure 5. Learning-related changes in Segment 2 spatiotemporal pattern similarity for hippocampus and PPA. For every pair of overlapping routes, Segment 1 corresponded to the paths that actually overlapped, whereas Segment 2 corresponded to the route-unique paths (after the routes diverged; Fig. 1a,b). (**a**) In hippocampus, Segment 2 overlapping route similarity did not change across learning relative to non-overlapping route similarity ($F_{1,39}$ = 0.34, P = 0.56). An overlap (overlapping/non-overlapping) x learning (1st half/2nd half) x segment (1st segment/2nd segment) ANOVA revealed a marginally significant three-way interaction ($F_{1,39}$ = 4.06, P = 0.0508), reflecting the relative selectivity of learning-related representational changes to Segment 1. The difference between same route and overlapping route similarity in hippocampus was not significant for Segment 2 in either the 1st half ($F_{1,39}$ = 0.05, P = 0.83) or 2nd half of learning ($F_{1,39}$ = 0.01, P = 0.94). (**b**) In the 2nd half of learning, the hippocampal reversal effect (overlapping route similarity < non-overlapping route similarity) was significant for Segment 1 ($F_{1,39}$ = 14.20, P = 0.0005), but not Segment 2 ($F_{1,39}$ = 0.31, P = 0.58) and the interaction between overlap and segment was significant ($F_{1,39}$ = 4.28, P = 0.045). In other words, by the end of learning, hippocampal representations of overlapping routes were relatively less similar during the portions of the routes that actually overlapped (**c**) TR-by-TR comparison of spatial pattern similarity for overlapping vs. non-overlapping routes. Transition TRs reflect time points that included the end of Segment 1 and the beginning of Segment 2. (**d**) Segment 2 data from PPA. In both learning halves, overlapping route similarity was significantly lower than same route similarity (1st half: $F_{1,39}$ = 7.44, P = 0.01; 2nd half: $F_{1,39}$ = 8.03, P = 0.007). Importantly, differences between same vs. overlapping routes were not significant for PPA in Segment 1 (Fig. 3d). In other words, PPA representations of overlapping routes were discriminable from non-overlapping routes only once the overlapping routes diverged. Notably, for Segment 2 data in PPA, there was no difference between overlapping and non-overlapping routes in either half (1st half: $F_{1,39}$ = 0.18, P = 0.68; 2nd half: $F_{1,39}$ = 0.32, P = 0.57). Thus, although PPA activity patterns diverged once overlapping routes diverged (Segment 2), there was no evidence for a reversal effect for Segment 2 data in PPA. Considering 2nd half data, there was a significant overlap (overlapping vs. non-overlapping) x segment (Segment 1 vs. Segment 2) x region (hippocampus vs. PPA) interaction ($F_{1,39}$ = 5.97, P = 0.019), reflecting the relative differences across regions in representational structure by the end of learning. Error bars reflect +/- SEM. P < .1, * P < 0.05, ** P < 0.01, *** P < 0.001



Supplementary Figure 6. Learning-related changes in PPA spatiotemporal pattern similarity (Segment 1 only) for each fMRI experiment. (**a**) In Experiment 1, the interaction between learning (1st vs. 2nd half) and overlap (overlapping vs. non-overlapping routes) was not significant ($F_{1,19}$ = 0.45, P = 0.51). In both the 1st and 2nd halves of learning, overlapping route similarity was significantly greater than non-overlapping route similarity (1st half: t_{19} = 4.56, P = 0.0002; 2nd half: t_{19} = 4.76, P = 0.0001). (**b**) In Experiment 2, the interaction between learning (1st vs. 2nd half) and overlap (overlapping vs. non-overlapping routes) was significant ($F_{1,20}$ = 5.19, P = 0.034), reflecting a relative decrease in overlapping route similarity across learning. Whereas overlapping route similarity was greater than non-overlapping route similarity in the 1st half of learning (t_{20} = 2.27, P = 0.034), there was no difference between overlapping and non-overlapping route similarity in the 2nd half of learning (t_{20} = 0.55, P = 0.59). Error bars reflect +/- SEM. * P < 0.05, ** P < 0.01, *** P < 0.001.



Supplementary Figure 7. Comparison of learning-related changes in spatiotemporal pattern similarity (Segment 1 only) for hippocampus vs. retrosplenial cortex (RSC). In addition to PPA (our primary control region), retrosplenial cortex (RSC) has also been implicated in scene processing, spatial navigation, and episodic memory retrieval[1-3]. Within RSC, overlapping route similarity decreased across learning, relative to non-overlapping route similarity ($F_{1,39}$ = 4.45, P = 0.041). In the 1st half of learning, overlapping route similarly was significantly greater than non-overlapping route similarity ($F_{1,39}$ = 16.10, P = 0.00026). In the 2nd half, however, there was no difference between overlapping and non-overlapping route similarity ($F_{1,39}$ = 0.11, P = 0.74). Thus, although overlapping route similarity decreased relative to non-overlapping route similarity, there was no evidence of the reversal effect (overlapping route similarity < non-overlapping route similarity) that was observed in the hippocampus. There was also no learning-related changes in overlapping route similarity relative to same route similarity ($F_{1,39}$ = 0.18 , P = 0.67). Overlapping route similarity did not differ from same route similarity in either the 1st half ($F_{1,39}$ = 1.41 , P = 0.24) or 2nd half of learning ($F_{1,39}$ = 1.91 ,P = 0.18). For comparison, data from the hippocampus are shown here (identical to Fig. 3c). The difference in representational structure across hippocampus and RSC was confirmed by a significant interaction between overlap (overlapping vs. non-overlapping) and region (hippocampus vs. RSC) within the 2nd half of learning ($F_{1,39}$ = 10.23, P = 0.0027). Thus, although RSC patterns did change with learning, the end states of learning were qualitatively different across hippocampus and RSC. Error bars reflect +/- SEM. * P < 0.05, ** P < 0.01, *** P < 0.001



Supplementary Figure 8. Hippocampal pattern similarity as a function of route destination. In Experiment 2, non-overlapping routes included pairs of routes that terminated at distinct destinations as well as pairs that terminated at the same destination (see Fig. 1b). To test whether hippocampal activity patterns reflected navigational goals[4] we compared Segment 1 spatiotemporal pattern similarity in the hippocampus across these two non-overlapping route conditions. If pattern similarity in the hippocampus reflects predicted destinations, then pattern similarity should be higher in the same destination condition than in the distinct destination condition. This difference in pattern similarity (same destination > distinct destination) would also be expected to increase over the course of learning as predictions improved. (**a**) Within Segment 1, there was no evidence for a learning-related change in similarity in the same destination condition relative to the distinct destination condition ($F_{1,20}$ = 0.53, P = 0.47). When considering the 2nd half data alone, there was no difference in similarity between the same and distinct destination conditions (t_{20} = 0.98, P = 0.34). These data indicate that spatiotemporal activity patterns elicited during Segment 1 did not reflect predicted destinations or navigational goals. (**b**) To test whether destination coding might occur at discrete time points within the route (e.g., at the beginning of each route) we compared the similarity between same destination and distinct destination on a TR-by-TR basis (using spatial pattern similarity). Data were separated by learning half (columns) and stimulus groups (rows). Data were divided by stimulus group because turns occurred at slightly different time points (TRs) across the stimulus groups. Pattern similarity did not significantly differ between the same and distinct route conditions at any time point across halves or stimulus groups (Ps > 0.05). Error bars reflect +/- SEM.

Supplement References

1. Vann, S. D., Aggleton, J. P. & Maguire, E. A. What does the retrosplenial cortex do? *Nature Reviews Neuroscience* **10,** 792–802 (2009).

2. Epstein, R. A. Parahippocampal and retrosplenial contributions to human spatial navigation. *Trends in Cognitive Sciences* **12,** 388–396 (2008).

3. Marchette, S. A., Vass, L. K., Ryan, J. & Epstein, R. A. Outside Looking In: Landmark Generalization in the Human Navigational System. *Journal of Neuroscience* **35,** 14896–14908 (2015).

4. Brown, T. I. *et al.* Prospective representation of navigational goals in the human hippocampus. *Science* **352,** 1323–1326 (2016).