The hippocampus encodes distances in multidimensional feature space

Stephanie Theves[1,4*], Guillén Fernandez[1], Christian F. Doeller [2,3*]

[1] Donders Institute for Brain, Cognition, and Behaviour, Radboud University and Radboud University Medical Center, Nijmegen, The Netherlands.

2 Kavli Institute for Systems Neuroscience, Centre for Neural Computation, The Egil and Pauline Braathen and Fred Kavli Centre for Cortical Microcircuits, NTNU, Norwegian University of Science and Technology, Trondheim, Norway

[3] Max-Planck-Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

[4] Lead Contact

*Correspondence: s.theves@donders.ru.nl, christian.doeller@donders.ru.nl

Keywords: cognitive geometry, cognitive map, conceptual knowledge, spatial coding

Summary

The hippocampal formation encodes maps of the physical environment [1, 2, 3, 4, 5]. A key question in neuroscience is whether its spatial coding principles also provide a universal metric for the organization of non-spatial information. Initial evidence comes from studies revealing directional modulation of fMRI responses in humans [6, 7] during navigation through abstract spaces and the involvement of place and grid cells in encoding of non-spatial feature dimensions [8]. However, a critical feature of a map-like representation is information about distances between locations, which has yet only been demonstrated for physical space [4, 9]. Here we probe whether the hippocampus similarly encodes distances between points in an abstract space spanned by continuous stimulus-feature dimensions that were relevant to the acquisition of a novel concept. We find that after learning, two-dimensional distances between individual positions in the abstract space were represented in the hippocampal multi-voxel pattern as well as in the univariate hippocampal signal as indexed by fMRI adaptation. These results support the notion that the hippocampus computes domain-general, multidimensional cognitive maps along continuous dimensions.

Results

In a feedback-based categorization task, participants acquired the concept of two abstract stimulus categories, which was defined within a two-dimensional space along the stimulus-dimensions of opacity and circle size with the diagonal as category boundary (Figure 1). Prior to the categorization task, six objects were associated with six specific abstract stimuli (randomized across participants), to place the objects in clearly defined distances to each other. Finally, the abstract space was explored in a free-recall task ('Navigation', Figure 1), which required collecting certain objects by editing the features of a random abstract stimulus until it matched the stimulus associated with the requested object. Learning took place over the course of two days. To ensure hippocampal dependency of the acquired information following a night of sleep on day 2, we introduced slight changes to the space by associating two of the objects with new abstract stimuli. All six associations were treated equally during learning on Day 2. Critically, fMRI data were acquired during object viewing blocks, immediately before and after the learning phase in order to test whether hippocampal responses to the objects correspond to their relative two-dimensional distances in the feature space.

Object detection task (in fMRI)

The six objects that were associated with abstract stimuli during learning, plus an additional catch object were presented multiple times in a randomized sequence that was identical between the pre- and the post-learning block. Participants indicated via button press, whether or not a presented object was the catch object. The task was performed with high accuracy (pre-learning (mean±SD): 97.19 ± 0.98%, post-learning: 98.01±1.74%), indicating that participants paid attention to the objects.

Learning tasks

The six object-abstract stimulus associations were learned with high accuracy rates (final block on day 1 (96±4%); final block on day 2 (83±2%)) (Figure S1 A). In the categorization task, each participant's performance exceeded chance-level (=50%) in all six analysis-bins a 60 trials and peaked with an average accuracy rate (across participants) of 80% in the final bin (Figure S1 B). When being asked for their categorization strategy at the end of the experiment, no participant indicated the use of a spatial rule or an imagination strategy of the spatial layout. Navigation/free recall-task data revealed that during the second half of the task, participants' paths (=feature editing process) from start-stimuli to stimuli associated with the target-objects were significantly more consistent with the shortest possible paths (day 1: $t(47)=8.1$, $p<.001$; Day 2: $t(23)= 2.0$, $p=.042$) (Figure S1 C). This shift to taking shorter paths through the feature space might reflect improved recall of associations as well as the build-up of a map-like mental representation.

Neuroimaging

We hypothesized that the hippocampus maps distances between objects in an abstract multidimensional space defined along the stimulus-feature dimensions, akin to physical space, as closer objects being represented more similar, akin to [4] and [9] in the spatial domain. First, we probed the representation of two-dimensional distances in the hippocampus in an fMRI adaptation analysis. To this end, we regressed each object's distance to the preceding object against each voxel's time series in the post-learning block and submitted the resulting beta weights, averaged across all voxels within the hippocampal ROI, to group level analysis. We found that hippocampal responses decreased with decreasing distance between objects (HC: $t(33)=1.943$, $p=.027$) (Figure 2). Hippocampal adaptation did not differ significantly across sub-regions (one-factorial permutation ANOVA $F=0.977$, $p=0.442$). In post-hoc analyses, we did not observe effects in other brain regions that survived correction for multiple comparisons on a whole-brain level (cluster-extend-based thresholding, $z=2.3$, $p=.05$). The distance-effect did not occur in regions in which we would not expect coding for distances between the objects in multidimensional feature space (postcentral gyrus: $p=0.301$, $t=0.525$), specifically also not in sites along the ventral visual stream (LOC: $p=0.106$, $t=1.300$) which have been shown to code for the two feature dimensions (luminance, size) alone [10] (Figure S2 A, B).

Additionally, we investigated representational changes as a consequence of the learning process in the hippocampal multivoxel-pattern. We tested if changes in hippocampal pattern similarity across objects from the pre-learning to the post-learning block corresponded to the distances between objects as introduced in the learning phase. The overall hippocampal pattern did not significantly reflect the two-dimensional distances (bilateral HC, $t(33)=-0.4331$, $p=0.3295$). However, in accordance with previous studies demonstrating coding differences along the long-axes as well as between hemispheres [11, 12], we observed a significant main effect of ROIs ($F=3.647$, $p=0.008$) in a one-factorial permutation ANOVA and post-hoc tests showed that the pattern in the anterior right HC encoded the two-dimensional distances (anterior right HC $t(33)=-2.869$, $p=.012$; anterior left HC: $t(33)=-0.173$, $p=.500$, posterior left HC: $t=1.593$, $p=.797$, posterior right HC: $t(33)=-.728$, $p=.460$). As expected, pattern similarity increased for small and decreased for large distances (Figure 3). We did not observe this effect in regions in which we would not expect coding for abstract distances (postcentral gyrus: $p=0.468$, $t=-0.091$; LOC: $p=0.766$, $t=0.677$) (Figure S2 A, B). These analyses show that the hippocampus represents the two-dimensional distances between objects located at different positions in abstract feature space.

Discussion

For the first time, we demonstrate that during concept learning the human hippocampus encodes distances in a multidimensional abstract space akin to distances in navigational space. After participants had encountered an abstract space that was defined along two task-relevant continuous perceptual stimulus-feature dimensions and had learned to associate objects with specific stimuli in this space, their hippocampal responses to the objects reflected the relative two-dimensional distances between the objects. First, hippocampal adaptation scaled with the distance between successively presented objects, and second, multivoxel pattern similarity across objects corresponded to their pairwise distances.

So far the idea that spatial coding principles apply also to non-spatial domains was supported by studies indicating directional modulation of fMRI responses [6, 7] and the involvement of spatially tuned cells in the rat hippocampus and entorhinal cortex [8] in representing non-spatial feature dimensions. [6] demonstrated that the BOLD signal in vmPFC and entorhinal cortex (EC) encoded the path angle during navigation in a two-dimensional space defined by the length of two stimulus features, while [7] showed that during a virtual role-playing game, the hippocampal response tracked the angle between a participant's viewpoint and a character's position in space defined by two social dimensions. Comparing the current paradigm to [6] it's important to note that the grid-cell like signal in EC and vmPFC was related the use of structural information only, while in the current paradigm mapping the objects according to their conceptual properties, requires a conjunction of structural information (two-dimensional space defined by the ratio of opacity and circle size) and object specific (sensory, semantic) information. We do not find evidence for a two-dimensional distance representation in the EC (mask from Jülich Histological atlas, thresholded at 50% probability; adaptation: p=0.064, t=1.538, RSA: p=0.769, t=0.740). This different localization of map-like representations of multidimensional non-spatial information in EC and hippocampus between the two studies fits well with the idea that the EC extracts the generalized structure of an environment while the hippocampus shows conjunctive codes to set specific information in temporal or spatial contexts. Furthermore, these studies focused on 'navigation- and direction-related' activity and it remained unclear if abstract information would also be represented in a map-like format independent of navigational demands (e.g. as in the present object viewing blocks). On that note, [13] showed that the entorhinal cortex maps associative strength of implicitly encoded (=not task-relevant), discrete relations, that were defined by the temporal distances between sequentially presented objects. While this finding provides evidence for the notion that the hippocampal formation encodes information along a non-spatial dimension independent of navigational demands, relations were defined via associative regularities of *discrete* variables rather than *abstracted* information that emerges from a combination of two *task-relevant continuous* dimensions as shown in the present study. Thus, our findings significantly go beyond previous research on navigation of abstract spaces [6, 7, 8] and distance coding in physical space [9, 4, 5], in support of the cognitive map theory [3] and ultimately in support of universal hippocampal coding principles underlying the formation of multidimensional spaces that are independent of cognitive domain.

Multi-feature similarity coding has also been investigated in the domain of visual perception [14]. Here, LOC was shown to encode conjoint representations of behaviorally integral, but not of separable feature dimensions during stimulus presentation. Congruent with [14] we do not find a combined representation of our two separable dimensions in LOC either. A critical difference of our paradigm though is the cognitive nature of the task, involving prior association of 'perceptually unrelated' objects to two task-relevant stimulus feature dimensions and the visual absence of these features during our critical fMRI measurement, imposing the question of how higher-level cognitive areas such as the hippocampus would treat two task-relevant but perceptually separable dimensions. In this regard, our results offer insight in potential differences in multi-feature similarity coding across cognitive processes and brain regions.

Importantly, the changes in neural object representations can only be attributed to the introduced distances between the objects, as the object-to-abstract stimulus assignment was randomized across participants. As during the object viewing blocks participants were not instructed to retrieve any information acquired in the learning phase, a fairly automatic process is conceivable.

The distance effect reflected in hippocampal fMRI adaptation could, in principle, indicate the imagination of the path through the two-dimensional feature space (=editing features in the free-recall task) between stimuli associated with the preceding and currently presented object, where longer paths would require more intense processing and result in less adaptation. We consider this unlikely, as participants were neither instructed to perform any cognitive operation involving distance information, nor were the paths between objects ever experienced directly. Instead, 'navigation' to objects only occurred from different start stimuli and distances would thus need to be inferred indirectly, e.g. via construction of a mental map. In this context, adaptation among neurons with partially overlapping place fields might account for the two-dimensional distance effect on hippocampal adaptation.

As in the present design perceptual and conceptual similarity map onto the same two-dimensional space, it is difficult to explicitly distinguish whether the hippocampus maps the two-dimensional distances between objects for the purpose of concept learning or merely in accordance with the perceptual similarity of their associated abstract stimuli. We consider the latter explanation less probable given that the hippocampal formation was shown to map continuous sensory dimensions only if they are behaviorally relevant [8] and behavioral relevance is in the present study to a significant degree ascribed via concept learning (the combination of values on both dimensions assigns the category to a given symbol). Still, further research is necessary to explicitly distinguish these options. One might further question whether the representation of distances between the objects could also be explained by perceptual similarity of the input to the hippocampus during a potential pattern completion process in which the associated abstract stimuli would be recalled. We regard this unlikely for two reasons: First, participants were not instructed to recall the associated stimulus during presentation of the objects in the scanner. As they were further engaged in an orthogonal object detection task, increasing their cognitive load, it is questionable

that participants spend additional resources on recalling currently irrelevant information. Second, if the hippocampal distance effects would reflect a two-dimensional perceptual similarity code forwarded to the hippocampus by upstream regions during a pattern completion process, we would expect the distance effects in the respective sensory representation cortex (LOC for processing of luminance and size [10]), which we did not observe in the present study. Together this suggests, that in the present study the hippocampus encodes a combined multidimensional representation of task-relevant non-spatial dimensions entailing relational distances across multiple entities in an abstract space for the purpose of concept learning.

We probed distance representations using both fMRI adaptation as well as representational similarity analysis. The representation of two-dimensional distances in the hippocampus after learning was shown in both analyses, which parallels previous demonstrations of spatial distance coding in the human hippocampus (adaptation: [4], similarity-based: [9]). The distance in abstract space between successively presented objects was reflected in the adaptation over all hippocampal voxels while the distance effect on multivoxel pattern similarity was located anteriorly. This regional difference between the analyses might be due to the anterior-posterior gradient in memory integration [15, 11] and the different degrees to which information was integrated in each analysis: While the representational similarity analysis reflects a multistep integration over all repetitions of each object, fMRI adaptation is a 1-step back analysis and thus is likely more sensitive to effects on a fine-grained scale. Further, it has been demonstrated that univariate and multi-voxel pattern analyses are susceptible to different sources of variance (subject-level variance in mean activation and voxel-level variability in the effect of a condition within subjects, respectively) and diverging results between the two do thus not per se allow conclusions about the nature of the neural code [16]. It might also be worth noting that in the present RSA as compared to the adaptation analysis we additionally remove noise on the single subject level by subtracting the pre-learning from the post-learning correlation matrix. While for both analyses the reported group-level effects can, due to object randomization, be ascribed to our critical manipulation, the higher susceptibility to noise on the individual subject level in the adaptation analysis might be one reason for lower effects as compared to the RSA results. While the relationship between the present distance code detected in the hippocampal univariate signal and multi-voxel pattern cannot be exactly determined, we can, importantly, confirm that both markers used to identify distance coding in spatial navigation (adaptation: [4], similarity-based: [9]) can be successfully applied to study abstract spaces.

The distance code for abstract information demonstrated here provides compelling evidence for the idea that the hippocampus encodes new information in form of a cognitive map [3]. Such a map-like cognitive representation allows inferences about relationships that are not directly experienced (i.e. participants never navigated directly between the six objects and the distances, reflected in hippocampal response pattern, might have been inferred through the creation of a two-dimensional representation defined along the two stimulus-feature dimensions) and therefore enables cognitive operations such as reasoning,

generalization and abstraction [17, 18, 19, 20] which are necessary for the organization and use of knowledge [21].

An open question refers to the generalizability of the map-like representation to more complex real world concepts which are often defined by multiple, not necessarily orthogonal dimensions. It remains to be tested whether hippocampal activity would reflect the distances between locations defined in a multidimensional space and if the scale of distance representations would be shaped by correlated dimensions.

Furthermore, the question emerges of how the category boundary in the present space would affect grid-coding during navigation as demonstrated for a homogeneous and continuous feature space [6]. In physical space, environmental boundaries were shown to distort [22, 23, 24] or segment grid patterns [25] in entorhinal cortex, while it was also shown that with navigational experience in two neighboring compartments, grids rescale to provide a metric for a unified large-scale environment [26, 27]. The latter effect could occur in the current concept learning paradigm, if coarse categorical information is acquired before representing the exact distances across all objects.

By demonstrating that the hippocampus encodes distances between locations in a multidimensional feature space, the present study critically supports the idea that hippocampal coding principles are independent of the cognitive domain and can provide a suitable format to represent conceptual knowledge.

Acknowledgements

CFD's research is supported by the Max Planck Society; the Kavli Foundation; the European Research Council (ERC-CoG GEOCOG 724836), the Centre of Excellence scheme of the Research Council of Norway – Centre for Neural Computation (223262/F50), The Egil and Pauline Braathen and Fred Kavli Centre for Cortical Microcircuits, the National Infrastructure scheme of the Research Council of Norway – NORBRAIN (197467/F50); and the Netherlands Organisation for Scientific Research (NWO-Vidi 452-12-009; NWO-Gravitation 024-001-006; NWO-MaGW 406-14-114; NWO-MaGW 406-15-291).

Author Contributions

S.T. and C.F.D. conceived the experiment. S.T. performed the experiment and analyzed the data. S.T. and C.F.D. wrote the manuscript. S.T., C.F.D., and G.F. discussed the results and contributed to the manuscript.

Declaration of Interests

The authors declare no competing interests.

References

1. O'Keefe, J. and Dostrovsky, J. (1971). The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. Brain Res. *34*(1), 171-5.

2. Hafting, T., Fyhn, M., Molden, S., Moser, M.B., Moser, E.I. (2005). Microstructure of a spatial map in the entorhinal cortex. Nature *436*(7052), 801-6.

3. O'Keefe, J. and Nadel, L. (1978). The Hippocampus as a Cognitive Map. Clarendon Press.

4. Morgan, L.K., Macevoy, S.P., Aguirre, G.K., Epstein, R.A. (2011). Distances between real-world locations are represented in the human hippocampus. J Neurosci *31*(4), 1238-45.

5. Howard, L.R., Javadi, A.H., Yu, Y., Mill, R.D., Morrison, L.C., Knight, R., Loftus, M.M., Staskute, L., Spiers, H.J. (2014). The hippocampus and entorhinal cortex encode the path and Euclidean distances to goals during navigation. Current Buiology *24*, 1331-1340.

6. Constantinescu, A.O., O'Reilly, J.X., Behrens, T.E.J. (2016). Organizing conceptual knowledge in humans with a gridlike code. Science *352*(6292), 1464-1468.

7. Tavares, R.M., Mendelsohn, A., Grossman, Y., Williams, C.H., Shapiro, M., Trope, Y., Schiller, D. (2015). A Map for Social Navigation in the Human Brain. Neuron *87*(1), 231-43.

8. Aronov, D., Nevers, R., Tank, D.W. (2017). Mapping of a non-spatial dimension by the hippocampal-entorhinal circuit. Nature *543*(7647), 719-722.

9. Deuker, L., Bellmund, J.L., Navarro Schröder, T., Doeller, C.F. (2016). An event map of memory space in the hippocampus. Elife *6*, 5.

10. Pinel, P., Piazza, M., Le Bihan, D., Dehaene, S. (2004). Distributed and overlapping cerebral representations of number, size, and luminance during comparative judgments. Neuron *41*, 983-993.

11. Collin, S.H., Milivojevic, B., Doeller, C.F. (2015). Memory hierarchies map onto the hippocampal long axis in humans. Nat Neurosci *18*(11), 1562-4.

12. Kühn, S. & Gallinat, J. (2014). Segregating cognitive functions within hippocampal formation: a quantitative meta-analysis on spatial navigation and episodic memory. Hum Brain Mapp *35*(4), 1129-42.

13. Garvert, M., Dolan, R.J., Behrens, T.E. (2017). A map of abstract relational knowledge in the human hippocampal-entorhinal cortex. Elife (6).

14. Drucker, D.M., Kerr, W.T., Aguirre, G.K. (2009). Distinguishing conjoint and independent neural tuning for stimulus features with fMRI adaptation. J Neurophysiology *101*(6), 3310-24.

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

16. Davis, T., LaRocque, K.F., Mumford, J., Norman, K.A., Wagner, A.D., Poldrack, R.A. (2014). What do differences between multi-voxel and univariate analysis mean? How subject-, voxel-, and trial-level variance impact fMRI Analysis. Neuroimage *97*, 271-283.

17. Hummel, J. E., & Holyoak, K. J. (2005). Relational Reasoning in a Neurally Plausible Cognitive Architecture. An Overview of the LISA Project. Current Directions in Psychological Science, 14(3), 153–157.

18. Rosch, E., "Principles of Categorization", pp. 27–48 in Rosch, E. & Lloyd, B.B. (eds), Cognition and Categorization, Lawrence Erlbaum Associates, Publishers, (Hillsdale), 1978.

19. Smith, E. E., & Medin, D. L. (1981). Categories and concepts. Cambridge, MA: Harvard University Press.

20. Skorstad, J., Gentner, D., & Medin, D. (1988). Abstraction processes during concept learning: A structural view. Proceedings of the 10th Annual Conference of the Cognitive Science Society, 419-425.

21. Kumaran, D., McClelland, JL. (2012). Generalization Through the Recurrent Interaction of Episodic Memories: A Model of the Hippocampal System. Psychol. Rev. *119*(3), 573-616.

22. Barry, C., Hayman, R., Burgess, N., and Jeffery, K.J. (2007). Experience-dependent rescaling of entorhinal grids. Nat. Neurosci. *10*, 682–684.

23. Stensola, T., Stensola, H., Moser, M.B., and Moser, E.I. (2015). Shearing-induced asymmetry in entorhinal grid cells. Nature *518*, 207–212.

24. Krupic, J., Bauza, M., Burton, S., Barry, C., and O'Keefe, J. (2015). Grid cell symmetry is shaped by environmental geometry. Nature *518*, 232–235.

25. Derdikman, D., Whitlock, J.R., Tsao, A., Fyhn, M., Hafting, T., Moser, M.B., and Moser, E.I. (2009). Fragmentation of grid cell maps in a multicompartment environment. Nat. Neurosci. *12*, 1325–1332.

26. Carpenter, F., Manson, D., Jeffery, K., Burgess, N., Barry, C. (2015). Grid cells form a global representation of connected environments. Curr Biol *25*(9), 1176-82.

27. Wernle, T., Waaga, T., Mørreaunet, M., Treves, A., Moser, M., Moser, E. I. (2017). Integration of grid maps in merged environments. Nat Neurosci *20*.

28. Fanselow, M.S. & Dong, H. (2010). Are the dorsal and ventral hippocampus functionally distinct structures? Neuron *65*, 7–19.

29. Groppe, D.M. (2010). One sample/paired samples permutation t-test with correction for multiple comparisons. Available: http://www.mathworks.com/matlabcentral/fileexchange/29782-one-sample-paired-samples-permutation-t-test-with-correction-for-multiple-comparisons/content/mult_comp_perm_t1.m. Accessed 2017 May.

30. Maldjian, J., Laurienti, P.J., Kraft, R., Burdette, J.H. (2003). An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets. Neuroimage *19*, 1233–1239.

31. Blair, R.C. & Karniski, W. (1993). An alternative method for significance testing of waveform difference potentials. Psychophysiology.

32. Smith SM1, Jenkinson M, Woolrich MW, Beckmann CF, Behrens TE, Johansen-Berg H, Bannister PR, De Luca M, Drobnjak I, Flitney DE et al. (2004). Advances in functional and structural MR image analysis and implementation as FSL. Neuroimaage *23* Suppl 1, S208-19.

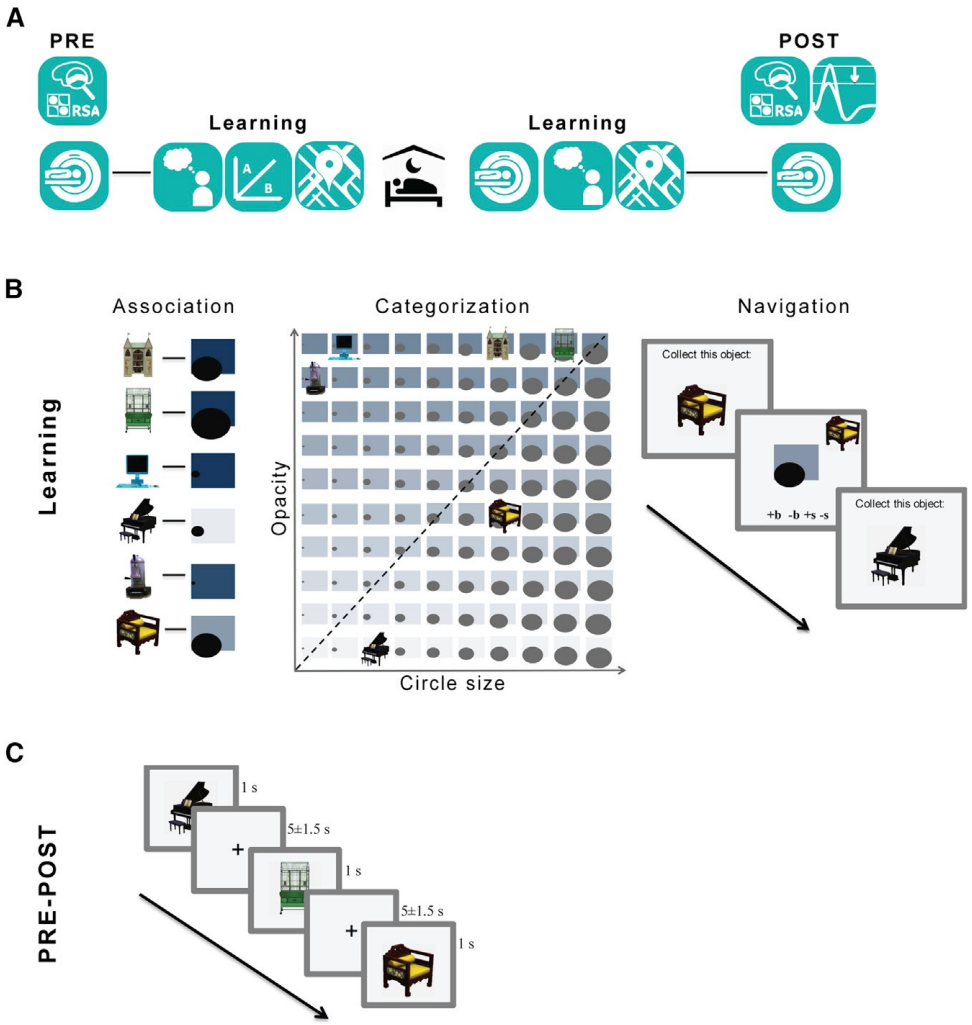


Figure 1. Two-dimensional concept learning tasks and scanning task.

(A) Experimental timeline. Learning (outside MRI; task-icons left to right: Association, categorization, navigation) took place over two days and was framed by object viewing blocks inside the scanner to measure the emergence of neural concept representations. B) *Left*: In the learning phase, participants associated six objects with six abstract stimuli (Association). *Middle:* Subsequently, they learned to categorize abstract stimuli into two categories (A and B) based on a relational rule (=concept) that set the diagonal through the two-dimensional feature space as category boundary (Categorization). *Right:* Participants 'navigated' the feature space by collecting requested objects through merging a 'start' stimulus into the stimulus associated with the object (Free recall/ Navigation). (C) In identical object viewing blocks inside the scanner, six different objects were presented in a randomized sequence with 30 repetitions per object. Objects were on screen for 1 s and were followed by a fixation cross for ITIs ranging from 3.5-6.5 s. Participants had to indicate via button press for each object whether it was the trampoline (catch-object) or any other object. Catch-trial rate was 10%. (See also Figure S1).

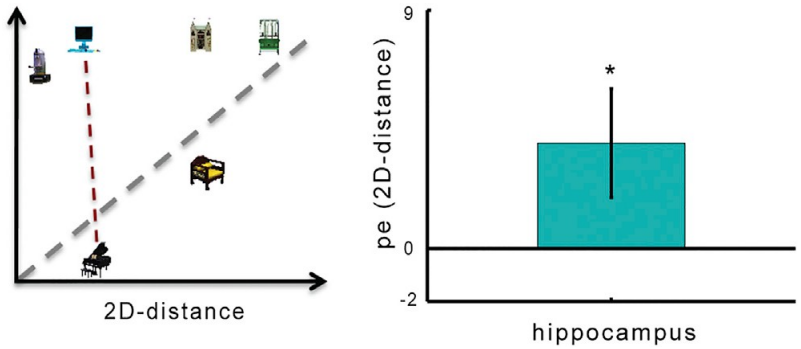


Figure 2. Two-dimensional hippocampal distance code for feature space revealed by BOLD adaptation.

Schematic of two-dimensional distances (red line) in feature space (left). Average of parameter estimates (pe) of the 'distance to preceding object' regressor in all hippocampal voxels. Hippocampal adaptation decreases with increasing two-dimensional distance between two successively presented objects (right). Asterisk (*) indicates significance at p=.05. Error bars indicate SEM. (See also Figure S2).

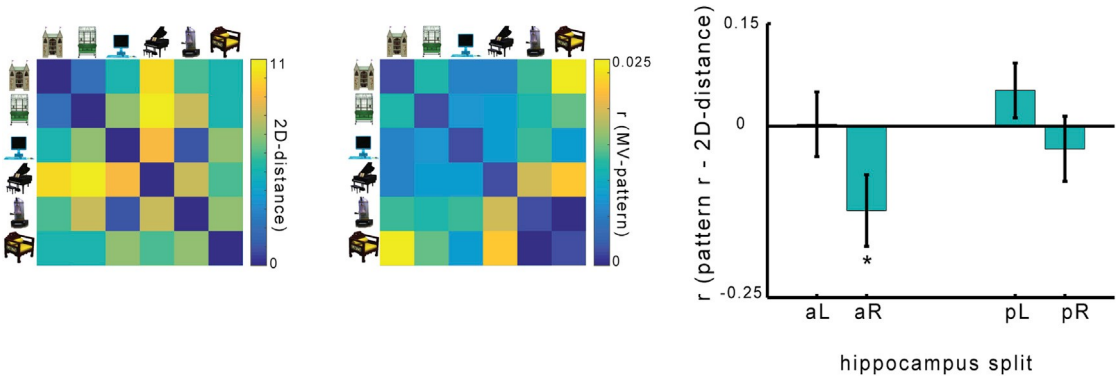


Figure 3. Hippocampal multi-voxel pattern reflects two-dimensional distances in feature space.

After learning relative to pre-learning baseline, two-dimensional distances between objects in feature space (left) significantly correlate with anterior right hippocampal pattern similarity across object pairs (middle; both matrices depict data of one example subject). Bars (right) depict the across-subject correlations between two-dimensional distances between objects and across-object pattern similarity in each ROI. Pattern similarity in the anterior right hippocampus increases significantly with decreasing distance. Asterisk (*) indicates significance at p=.05 corrected for multiple comparisons. Error bars indicate SEM. (See also Figure S2).

STAR★Methods

Contact for Reagent or Resource Sharing

Further information and requests for resources should be directed to Stephanie Theves (s.theves@donders.ru.nl).

Experimental Model and Subject Details

Thirty-six healthy students from the Radboud University campus participated in this study. All participants were right-handed and had normal or corrected-to-normal vision. Two participants were excluded from further analyses: One participant was excluded due to misunderstanding of the task instruction during the object-detection task and one due to technical problems during data acquisition. The final group consisted of 34 participants (24 females, age 18-30, mean= 23.6, SD= 2.9). All participants gave written informed consent and were financially compensated for participation. The study was approved by the local ethics committee (CMO Arnhem-Nijmegen, the Netherlands).

Method Details

Behavioral procedures

The study took place in two sessions over the course of two consecutive days, with a maximum of 30 hours between both sessions. A pre-learning object-viewing block (scanned) preceded the first learning phase (outside scanner). On the subsequent day a second learning phase was performed in between two additional object viewing blocks (scanned) identical to the first one (Figure 1A). During these blocks participants performed an object detection task (Figure 1C). The research question of this report is addressed by the analysis of the first (pre-learning) and the last (post-learning) object-viewing block and the report in the main text is thus restricted to those. For reasons of transparency, data of the middle scanning session are shown in Figure S2 C.

Object detection task: Images of seven objects (generated with the Video game Sims; www.thesims3.com), of which six where used in the following learning phases and one served as a catch-trial object, were presented in a pseudo-randomized sequence with a stimulus duration of 1 s and inter-stimulus-intervals of 3.5, 5, and 6.5 s (33,3% each). Participants were instructed to indicate for each object whether it is a trampoline (=catch trial object) or not, using a button box (buttons counterbalanced across participants). The task included 180 trials with a catch trial rate of 10%. Each object was presented equally often.

In the first learning phase, participants acquired a novel concept, defined by a two-dimensional stimulus-feature space (see '*Categorization'* and Figure 1B) as well as six associations between feature-space stimuli and the objects presented during the object-viewing blocks. Object to abstract stimulus assignment was randomized across participants, to ensure that similarities between neural object representations result from their relative distances in the abstract space but not from visual similarities. As on day 2, conceptual knowledge acquired on day 1 might have already entered the consolidation process, we introduced slight changes in the second learning phase to ensure hippocampal dependency of the acquired concept at the time of measurement (final fMRI session). Therefore, two of the six objects were associated with new abstract stimuli, thereby changing their positions in abstract space and the relative distances between the six objects. We acquired fMRI data in a middle scanning session as a backup at the beginning of day 1, in case participant's behavioral performance would be significantly de deteriorated by the updating manipulation. As his was not the case, we could restrict our analysis of the role of the hippocampus in mapping abstract distances to the final block which immediately followed acquisition and not a possible consolidation phase (middle scanning session). The first learning phase comprised an associative learning, categorization and navigation task, while in the second learning phase the associative learning and navigation task were performed.

Associative learning: Six object-stimulus associations had to be learned in seven alternating encoding- and test-blocks. In the encoding blocks, objects were presented next to their corresponding abstract stimulus and participants were instructed to memorize the presented pairs. The presentation order of the six pairs was pseudo-randomized with each object/stimulus being equally often presented on the left/right position of the screen. Pairs were presented for 2 s on the screen and each pair was shown 4x in the first encoding block and twice in the following encoding blocks (18 repetitions in total). Each encoding block was followed by a test block in which the object is presented in the center of the screen along with the six abstract stimuli displayed (in a randomized order) below the object. Every association was tested once per block in a randomized order. Participants selected the abstract stimulus associated with the presented object via key press (1-6) and received feedback (500 ms) on whether the choice was correct. Subsequent to the categorization task, associations were tested once more (14 tests per association with feedback; named 'final' in Figure S1 A). In the second learning phase, new associations (two objects were 'remapped' by assigning them to new stimuli) were acquired in seven alternating encoding- and test blocks with two repetitions per encoding block, once before and once after (summarized as 'final' in Figure S1 A) the navigation task, respectively. All six object-stimulus associations were included and treated the same way.

Categorization: Participants were instructed to categorize abstract stimuli (a black circle on a blue square) which varied along two independent feature dimensions (opacity of the blue feature and size of the circular feature, see Figure 1), into one of two categories (A- and B-symbols). Categories were, unknown

to the participant, delineated via the diagonal through the two-dimensional feature space (dashed line in Figure 1 B middle). Consequently, the ratio between opacity and circle size of a given stimulus defined whether it is an A- or B-symbol. Both features could take values from 1-10 (opacity: 10%-100% scale with 10% increase per step; circle size: decrease in size by 10% per each step with the biggest circle being 0.67''x 0,83''), resulting in 100 possible stimuli to sample the space. The 90 off-diagonal stimuli were presented 4 times in a randomized sequence. In each trial one abstract stimulus was presented in the center of the screen and its category had to be selected via keys press. Participants were given a maximum of 6 s to respond and each response was followed by feedback (500 ms). Instructions did not include any indications of a spatial rule. All trials were performed in a continuous stream. To depict the learning process, performance is plotted as the average in bins of 60 trials in Figure S1. Prior to the feedback-based categorization task, participants were given a 3-minutes exploration phase, during which they could freely up- and downregulate opacity and size of a given stimulus (using 4 adjacent keys; identical to the navigation task) while the corresponding category membership ('A' or 'B') was updated with every edition and displayed above the abstract stimulus (compare trial sequence 'navigation' in Figure 1B right: Instead of an object requested during navigation, an 'A' or 'B' is displayed.). This was done to accelerate learning. Subsequent to all experimental sessions, participants were asked to describe their categorization strategy, to interrogate whether they had a spatial rule or spatial layout in mind.

'Navigation'/Free recall: In each trial one of the six objects was presented followed by a random stimulus (selected from the total pool of 100 possible samples). Participants were instructed to 'collect' the displayed object by editing the features of the stimulus (again using 4 adjacent keys) until they match the stimulus associated with the object. A trial ended when the matching stimulus was created. Participants performed 96 trials in the first and 48 trials in the second learning phase. Along with strengthening participants' memory of the six associations through this free-recall, the task was supposed to familiarize participants with the conceptual context of the objects. Importantly, no distance relationships between the objects were introduced through this process, because participants did not navigate between stimuli associated with the objects, but instead started from random positions in the feature space. Furthermore, analyzing the length of the paths (=number of clicks made in the editing process) participants take provides a behavioral indication of a map-like representation.
All tasks were conducted using Presentation 16.4 (NBS), except the Navigation/Free-recall task, which was programmed using Anaconda 2.7 (Python).

MRI Methods

All images were acquired using a 3T PrismaFit MR scanner equipped with a 32 channel head coil (Siemens, Erlangen, Germany). A 4D multiband sequence (84 slices (multi-slice mode, interleaved), voxel size 2 mm isotropic, TR = 1500 ms, TE = 28 ms, flip angle = 65 deg, acceleration factor PE= 2, FOV =

210 mm) was used for functional image acquisition. In addition, a structural T1 sequence (MPRAGE, 1mm isotropic, TE = 3.03 ms, TR = 2,300 ms, flip angle = 8 deg, FOV = 256 × 256 × 192 mm) was acquired. Separate magnitude and phase images were used to create a gradient field map to correct for distortions (multiband sequence with voxel size of 3.5 × 3.5 × 2.0 mm, TR = 1,020 ms, TE=10 ms, flip angle = 45 deg).

Preprocessing of functional images was performed with FSL 5.0.9. Motion correction, high pass filtering at 100 s and distortion correction was applied to the functional data sets. (Exclusion criteria for excessive motion: Mean absolute displacement >2 mm or peak in absolute displacement >3.9 mm; mean and STD of absolute displacement of analyzed sample (mean±std): 0.349±0.157mm (pre) and 0.347±0.197mm (post). Spatial smoothing was only performed before the univariate analysis of the data. The data were not spatially smoothed before being subjected to representational similarity analysis (RSA), see below. The FSL brain extraction toolbox was used to create a skull-stripped structural image. The structural scans were down-sampled to 2 mm (matching the functional image resolution) and segmented into gray matter (GM), white matter (WM) and cerebro-spinal fluid (CSF). Mean-intensity values at each time point were extracted for WM and used as nuisance regressors in the general linear model (GLM) analyses (see below). Structural images were registered to the MNI template. For each functional dataset (pre-, post-learning, post-relearning) the preprocessed mean image was registered to the individual structural scan and the MNI template. The co-registration parameters of the mean functional image were applied to all functional volumes.

Quantification and Statistical Analysis

Univariate analysis

First level GLM: The two-dimensional distance between objects in the feature space was modelled in a GLM using a stimulus onset regressor indicating the occurrence of an object on the screen and a second regressor being parametrically weighted by the two-dimensional Euclidean distance between an object to its preceding object. Distances between positions in abstract space were calculated given the positions feature-based coordinates (values from 1-10 on the opacity and circle size dimension respectively. Step-sizes reflected 10% increases in opacity and circle size). Smaller distances were expected to result in lower signals, reflecting fMRI adaptation. Further, the GLM included regressors accounting for catch trials and button presses as well as six motion parameters as covariates. Resulting beta-maps were transformed to MNI space to extract the average beta value of each ROI for subsequent analysis. *Group-level analysis*: First-level beta estimates of the parametric distance regressor were averaged across all voxels within an ROI (Hippocampus: Figure 2; EC, LOC, PoCG: Figure S2) for each participant and the distribution of these values was tested for significance (at alpha=5%) using a one-sample permutation t-test [29] in which 1000 random permutations were computed to estimate the distribution of the null

hypothesis. To test for effects on the whole-brain level, individual contrasts of the parametric distance regressor were subjected to the second level analysis. Cluster extend-based thresholding (z=2.3, p=.05) was performed to correct for multiple comparisons.

Representational similarity analysis

First level GLM: Object-specific activity during stimulus-viewing blocks (pre- and post-learning) was modelled in a GLM with a stimulus onset regressor for each of the six objects. Furthermore, regressors accounting for the presence of a catch trial and the button press, as well as six motion parameters were included as covariates. Resulting beta-maps were transformed in MNI space before extracting multivoxel-patterns of each ROI for the next analysis step. *First level analysis:* We defined a priori regions of interest (ROIs, see below) and examined the correlation between across-voxel activation patterns of first-level beta estimates within these ROIs as a proxy of neural similarity. The six object regressors were considered as the regressors of interest, leading to a 6 x 6 matrix of correlations (Spearman's correlation coefficient). Each ROI's pre-learning object-to-object similarity matrix was subtracted from the post-learning matrix and the resulting neural change matrix was correlated (Spearman's correlation coefficient) with a 6 x 6 prediction matrix including the Euclidean distances (calculated as described in 'univariate analysis') between each object pair. Distance representations were assumed to be reflected in a negative correlation between distance and neural similarity.
Group-level analysis: The distribution of correlation coefficients was tested for significance (alpha=5%) across participants for each ROI (Hippocampus: Figure 3; LOC, PoCG: Figure S2) using one-sample permutation t-test [29] in which 1000 permutations were computed via random sign flip to estimate the distribution of the null hypothesis. P-values were corrected for multiple comparisons using the "tmax" method [31].
Due to clear directed predictions on the relations between neural pattern similarity/BOLD response and distance (e.g. increasing distance was supposed to be reflected in a decreased pattern similarity and fMRI adaptation) one-sided tests were applied.

ROI definition

Based on our a-priori hypotheses, analyses were restricted to the hippocampus
and a hippocampal mask was constructed using the WFU pickatlas [30]. In order to be sensitive to potential coding differences between anterior and posterior divisions of the hippocampus [28, 11], we split the hippocampal mask in approximately equally long parts along the long axis (posterior portion of the hippocampus: from Y = −40 to −30; mid-portion of the hippocampus: from Y = −29 to −19; anterior portion of the hippocampus: from Y = −18 to −4) for the left and right side, following [11], and selected the

resulting anterior and posterior portions as ROIs. Control regions probability maps of LOC, and PoCG from Harvard-Oxford Structural cortical atlas were thresholded at 50% probability.

Data and code availability

The data that support the findings of this study and the analysis code are available from the corresponding author upon request.

Key Resources Table

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and Algorithms		
MATLAB R2014A	MathWorks	https://www.mathworks.com
FSL 5.0.9	FMRIB; [33]	https://fsl.fmrib.ox.ac.uk/fsl/fslwiki
Presentation 16.4	Neurobehavioral Systems	https://www.neurobs.com
Anaconda 2.7	Python	https://anaconda.org/anaconda/python

Supplementary Information

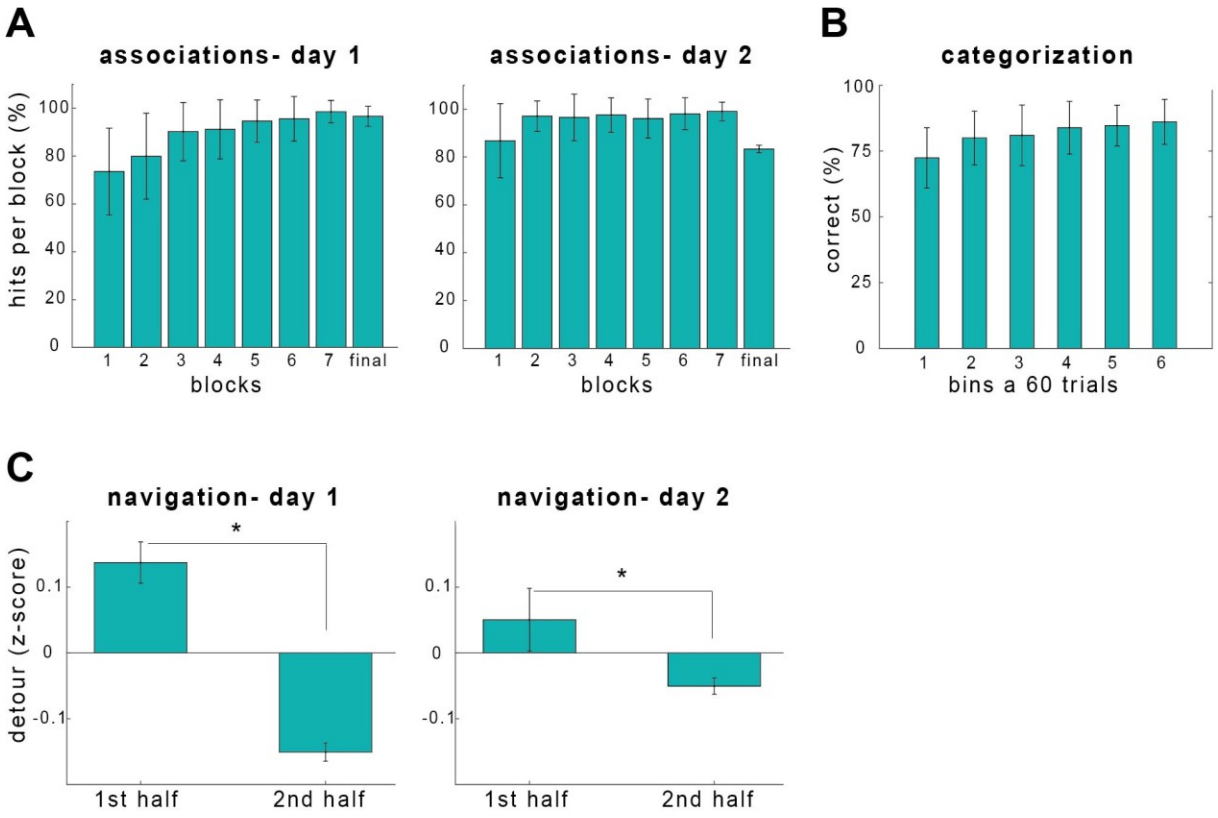
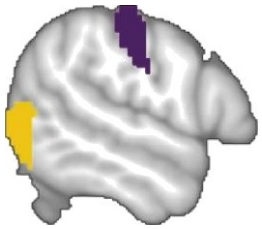


Figure S1. Participants learn to 'navigate' the concept space, Related to Figure 1.
(A) Percentage of correct responses in the associative learning task on Day 1 (left) and Day 2 (right), plotted for each block of the first session ('1-7') and over all trials of the second session ('final') of the respective day. (B) Percentage correct per analysis bin of 60 trials across participants in the categorization task. (C) Deviation of participants' paths (z-transformed) from shortest possible route (=number of required edits to transform the stimulus) during the 1st and 2nd half of the navigation task on Day 1 (left) and Day 2 (right), respectively. The paths that had to be navigated in both halves of the task were identical, but randomized in their order. On both days, participants' paths travelled in the 2nd half deviated significantly less from direct routes as compared to the 1st half. Error bars indicate STD.

A

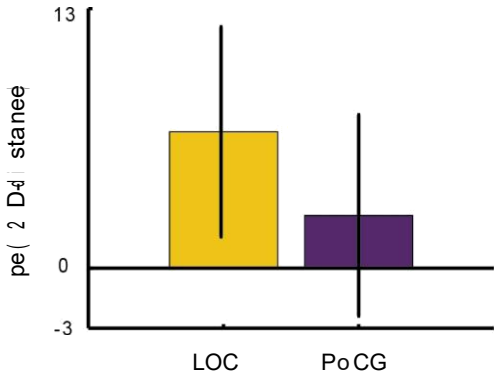


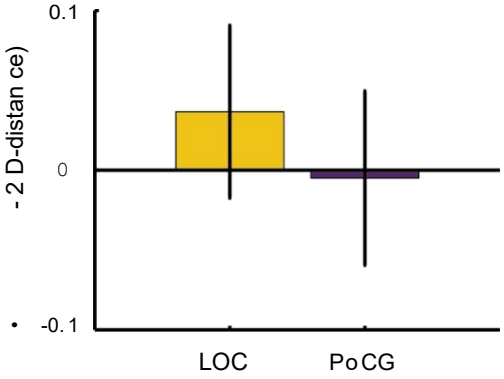
L ate ra l o cci pita l Cortex

Postcentra l Gyrus

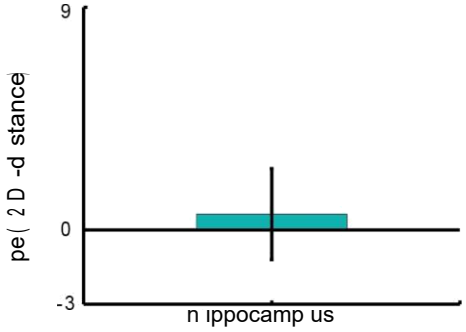
MNI: -55/-13/49

B



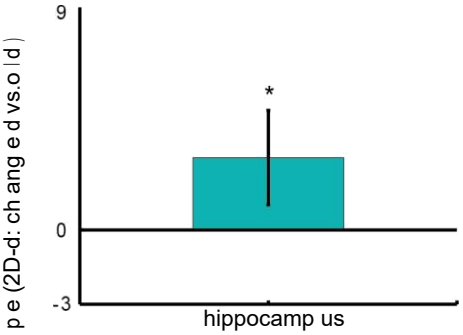


C



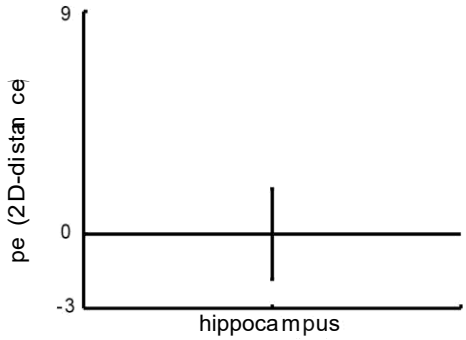
pre -l earnin g bloc k

D

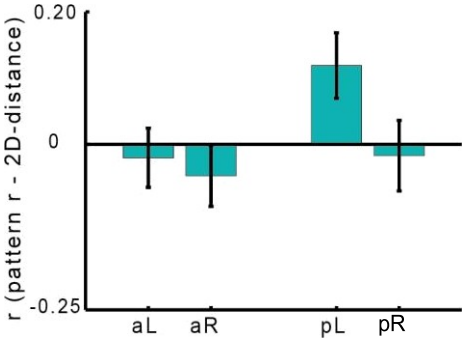


post-learn ing bl ock

E



middl e blo ck



mid dle blo ck

Figure S2. Distance effects on adaptation and pattern similarity in control regions and time-points, Related to Figures 2 and 3.

(A) Lateral Occipital Cortex (LOC) and Postcentral Gyrus (PoCG) defined by probability masks from Harvard-Oxford Structural Atlas, thresholded at 50% probability. (B) Parameter estimates of the 2D-distance regressor averaged over all voxels within an ROI (left) and correlation between 2D-distances across objects and across-object multivoxel pattern similarity in a given ROI (right). Two-dimensional distances were not encoded in LOC (adaptation: p=0.106, t=1.300, RSA: p=0.766, t=0.677) or PoCG (adaptation: p=0.301, t=0.525, RSA: p=0.468, t=-0.091) where we would not expect a distance code for abstract information. (C) Parameter estimates of 2D-distance regressor averaged across hippocampal voxels in the pre-learning session. Hippocampal adaptation does not scale with two-dimensional distances between the objects in the pre-learning fMRI session (p=0.368, t=0.347). (D) Contrast of parameter estimates of two separate regressors modelling changed distances (transitions between objects of which one changed 'position' in the second learning phase) vs. unchanged distances within the same GLM. The adaptation effect (representation of distances) was stronger for changed vs. unchanged distances ((*) approaching significance at p=0.0525, t=1.532), suggesting hippocampal dependency through gradual reactivation. (E) Two-dimensional distance effect on fMRI adaptation (left) and pattern similarity (right) in the middle scanning session at the beginning of day 2 prior to the second learning phase (adaptation: p=0.9, t=-0.00009; RSA: p=0.298, t=-0.511). Error bars indicate SEM.