
Responses to Reviewer #1

The relations between the two levels is interesting indeed especially with respect to the neuroanatomy, yet this reader is not convinced that this is what the data show and that the conclusion is right. Moreover results are poor and should be elaborated to support and extend the research claim

Reviewer #1 :

In this paper the authors re-analyze their 2017 impressive, yet not free of problems and debates, dataset as per semantic distance to inquire how are the "semantic space" created and the original "learned space" related to one another? while the second space was identified on the EC the semantic space was found in the hippocampus

We thank the reviewer for their helpful feedback. We have addressed their comments below point by point.

Although fmri adaptation is used both for link distance and semantic distance it is only the first which was designed with a pseudo random walker, while this is not supposed to influence a mere application of fmri adaptation, one cannot exclude that this design nonetheless influences the distinction between the two. This may account for the contradiction between these results and previous results.

We thank the reviewer for their comment. Although the original study was not designed to test the semantic effect, it is unlikely that this leads to a confound in the observed difference between link distance and semantic distance effects. While the pseudo-random walk participants were trained with was defined by a hidden graph with various link distance, the sequences presented to the participants at the point of testing (i.e. when fMRI adaptation was measured) was fully randomized (i.e. uncorrelated to either link distance or semantic distance). In fact, the sequence

of objects was generated such that each object followed each other object with equal probabilities. Moreover, the semantic distance and the link distance are not correlated in the current study (Spearman's Rho mean = .03, SD = .12, range = -.25 – .30, $t_{22} = 1.04$, $p = .31$). In other words, if we were to design the study anew to test the semantic distance, we would have set it up the same way as it is now. We therefore believe that it is unlikely that the specific sequence of events participants were exposed to in the scanner lead to the observed effect.

The main figure (Fig 3) is very pixelated and represented in slices, and not very informative. On the basic level, I suggest adapting it to the state of the art. It is btw a good practice to add subtitles within the figure. Moreover, Figure 3 is actually the only result that holds the authors claim. As such one would expect it to be more detailed and to be followed by additional analyses both to support the finding and to extend it or to show its implication to the field of cognitive neuroscience. However, none of these exist and the paper is therefore very poor. At this stage it is sufficient to hold as preliminary results for grant proposal, or as an introductory part to a paper but I am not sure how this limited and problematic finding may carry a full experimental paper in the Journal of Neuroscience.

We thank the reviewer for their comment. We now provided 300 dpi high-resolution figures in the revision following the journal's guideline to ensure that the image quality itself is not an issue.

However, we believe that the reviewer might be referring to the dual-coding plots used to display our fMRI results. We chose to display our results using the dual-coding plots, because there has been consensus in the literature that this represents the state of the art (Allen, Erhardt, & Calhoun, 2012; Zandbelt, 2017). Unlike standard fMRI displays, it simultaneously displays contrast estimate and *unthresholded* t-statistic maps, therefore providing unbiased and complete information about the results. Many recent papers in high-impact journals have employed this data presentation style (e.g., Ferrari et al., 2022, J. Neurosci; Richter & de Lange, 2019, eLife; van den Bosch et al., 2022, Nat. Commun.). We have additionally tuned a few parameters (e.g., opacity, line width) to improve the quality of these dual-coding plots and the updated figures are provided in the manuscript (also see below).

To provide a complete image of the distribution of semantic vs. statistical representations in cortical brain areas, we now present the link distance effect and semantic distance effect in equally spaced slices across the brain (Figure R1-1). As can be seen in Figure R1-1, both the link distance effect and the semantic distance effect are quite precisely located in the hippocampal formation (for more details see also whole-brain maps shared via the Donders repository: <https://doi.org/10.34973/8m6q-qj39>). While small clusters can also be observed in other regions, these do not survive correction for multiple comparisons. This indicates that the reported semantic effect is specific to the hippocampal formation. We discuss it in relation to the task relevance (see below). This plot has now been added to Figure 3 in the manuscript.

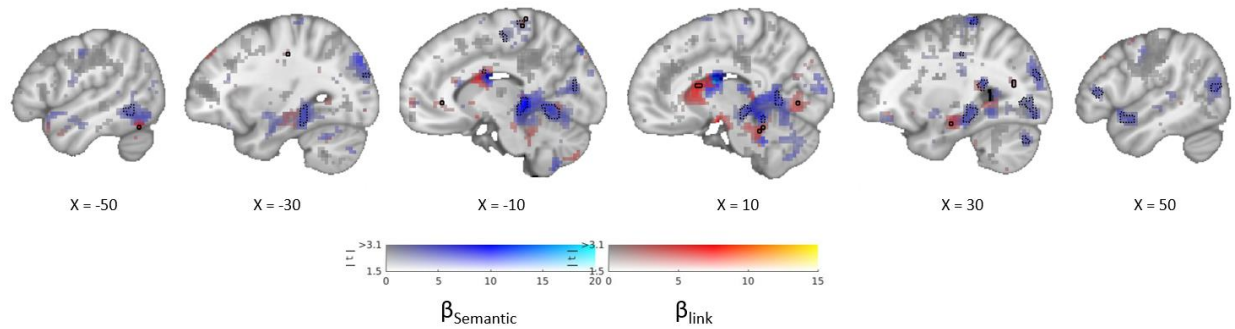


Figure R1-1. Link distance effect (red) and semantic distance effect (blue) presented at equally spaced slices across the cortex. Whole-brain results are displayed using Slice Display (Zandbelt, 2017) which takes a dual-coding data visualization approach (Allen, Erhardt, & Calhoun, 2012), with color indicating the parameter estimates and opacity the associated t statistics. Solid and dotted contours outline voxels exceeding the cluster-defining threshold of $p < 0.01$ uncorrected for the link and the semantic effects, respectively.

We have also performed several additional analyses to further support our findings and to connect to the rest of literature. First, we inspected the semantic distance effect specifically in two cortical areas which have previously been strongly associated with semantic processing: the left anterior temporal lobe (ATL, see Visser et al., 2010, *J. Cogn. Neurosci.* for a meta-analysis)

and the left angular gyrus (see Humphreys et al., 2021, TINS for a recent review). ATL in particular is thought to play a domain-general role in retrieving semantic knowledge (Gorno-Tempini and Price, 2001; Schneider et al., 2018). Additional analysis with small volume correction using these anatomically defined masks yield no suprathreshold clusters for either the link distance or the semantic distance effect (Figure R1-2). This provides strong evidence that the reported semantic effect is specific to the hippocampal formation. In the discussion, we examined how this relates to task relevance (see below).

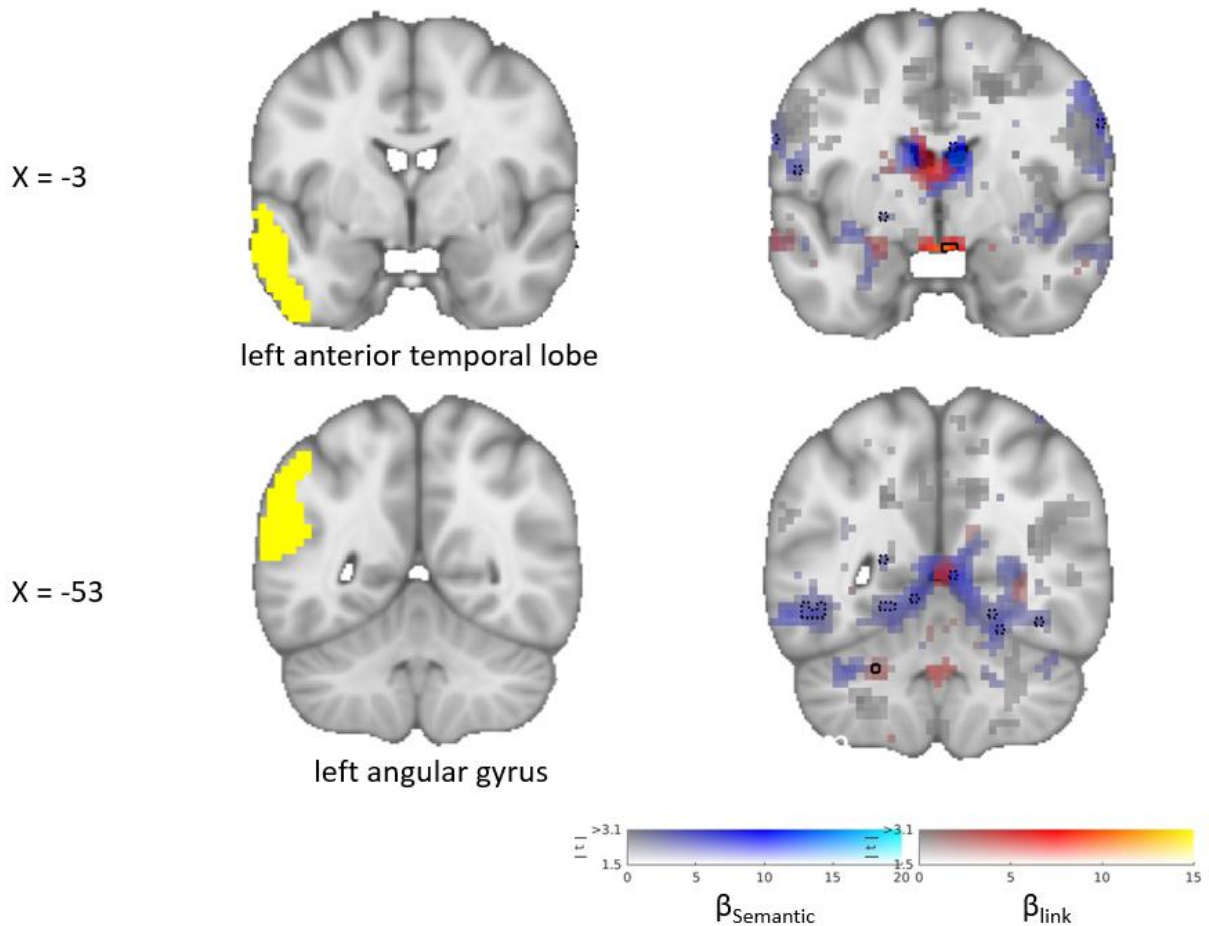


Figure R1-2. Two semantic-related ROIs used in the analysis (left panel) and corresponding data (right panel) representing link distance effect (red) and semantic distance effect (blue). Whole-brain results are displayed using Slice Display (Zandbelt, 2017) which takes a dual-coding data visualization approach (Allen, Erhardt, & Calhoun, 2012), with color indicating the parameter estimates and opacity the associated t statistics. Solid and dotted contours outline

voxels exceeding the threshold of $p < 0.01$ uncorrected for the link and the semantic effects, respectively.

We also added this additional analysis to the manuscript.

“To explore the cortical semantic representation, we performed additional SVC using two anatomically defined masks: the left anterior temporal lobe and the left angular gyrus, two regions previously reported to be important to semantic processing (Visser et al., 2010; Humphreys et al., 2021).” (in Methods, pages 9-10)

“No cortical regions showed fMRI adaptation effects as a function of semantic distance (all $ps > .71$, FWE corrected on the cluster level, Figure 3E), including specific regions of interest which have previously been associated with semantic processing: the left anterior temporal lobe (ATL, see Visser et al., 2010, for a meta-analysis) and the left angular gyrus (see Humphreys et al., 2021, for a recent review). ATL in particular is thought to play a domain-general role in retrieving semantic knowledge (Gorno-Tempini & Price, 2001; Schneider et al., 2018). Nonetheless, additional analyses with SVC using these anatomically defined masks yielded no suprathreshold clusters for either the link distance or the semantic distance effect, indicating that the reported semantic effect is specific to the hippocampal formation.” (in Results, pages 13-14)

References:

Visser, M., Jefferies, E., & Lambon Ralph, M. A. (2010). Semantic processing in the anterior temporal lobes: a meta-analysis of the functional neuroimaging literature. *Journal of Cognitive Neuroscience*, 22(6), 1083-1094. <https://doi.org/10.1162/jocn.2009.21309>

Humphreys, G. F., Lambon Ralph, M. A., & Simons, J. S. (2021). A unifying account of angular gyrus contributions to episodic and semantic cognition. *Trends in Neurosciences*, 44(6), 452-463. <https://doi.org/10.1016/j.tins.2021.01.006>

Gorno-Tempini, M. L., & Price, C. J. (2001). Identification of famous faces and buildings: a functional neuroimaging study of semantically unique items. *Brain*, 124(10), 2087-2097. <https://doi.org/10.1093/brain/124.10.2087>

Schneider, B., Heskje, J., Bruss, J., Tranel, D., & Belfi, A. M. (2018). The left temporal pole is a convergence region mediating the relation between names and semantic knowledge for unique entities: Further evidence from a “recognition-from-name” study in neurological patients. *Cortex*, 109, 14-24. <https://doi.org/10.1016/j.cortex.2018.08.026>

Second, to provide more details regarding our main results on the semantic vs. statistical representations in the hippocampal formation, we additionally performed a principal component analysis (PCA) to explore if there is a superior-inferior or an anterior-posterior gradient transiting from one effect to the other. Specifically, we extracted the group-level t-statistics for both the statistical and the semantic effects using an anatomically defined ROI combining the hippocampal formation (incl. hippocampus, entorhinal cortex, subiculum). We vectorized the voxel-based data and performed PCA on the link distance effect and the semantic effect. The analysis yielded two principal components, with the first component explaining 73% of the variance and capturing a negative correlation between statistical and semantic representations – consistent with what we have reported in the manuscript. We projected the loadings of the first component back into the voxel space (Figure R1-3). As we can see from Figure R1-3, there seems to be a clear posterior-superior to anterior-inferior gradient in the right hemisphere, which comes out a bit less strongly on the left side. However, the figure can also be read in line with the alternative interpretation of two separate clusters which reside predominantly in the hippocampus and in the entorhinal cortex, respectively – as argued in the original manuscript. There might be neurons that separately code for semantic or statistical information, and the linear change of loadings in the voxels might just be a result of averaging different types of neurons. Given the nature of the current fMRI data, it is hard to fully disentangle these two alternative hypotheses.

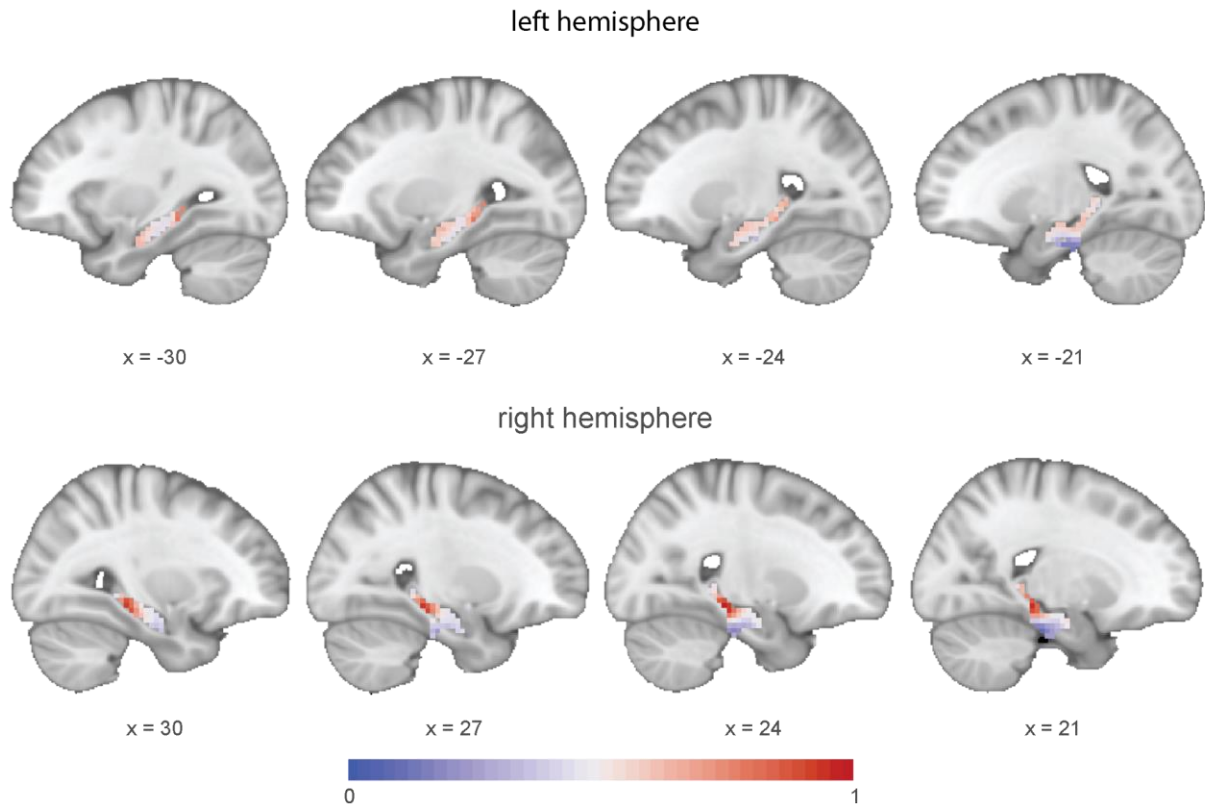


Figure R1-3. The first principal component from a principal component analysis of the statistical representation and the semantic presentation. The analysis is performed within the hippocampal formation (incl. hippocampus, entorhinal cortex, subiculum). A high loading means semantic representation is dominant, whereas a low loading means statistical representation is dominant. The loadings are normalized for visualization.

We now added this analysis to the manuscript. Figure R1-3 is included as a new Figure 4 in the manuscript.

“To further explore the spatial relationship between the link distance effect and the semantic distance effect, we sought to visualize their relationship in a hippocampal ROI that we defined anatomically (ie., same hippocampal mask used for SVC, combining hippocampus, entorhinal cortex, and subiculum). To this end, we conducted a principal component analysis (PCA) using group-level statistics for both the link and semantic distance effects. We extracted the group-level t-statistics for each effect using the anatomically defined hippocampal ROI. The voxel-based data was then vectorized and PCA was performed on the two distance effects. The

resulting principal components were normalized and projected back into the voxel space for visualization.” (in Methods, page 10)

“To better understand the spatial relationship between the link distance effect and the semantic distance effect, we visualized their relationship within an anatomically defined hippocampal ROI. Specifically, we performed a PCA analysis of both effects within the ROI. This analysis yielded two principal components, with the first explaining 73% of the variance and capturing a negative correlation between statistical and semantic representations. As shown in Figure 4, there appears to be a posterior-superior to anterior-inferior gradient in the right hemisphere, is less pronounced in the left hemisphere.” (in Results, page 14)

“Although these two maps seem to be anatomically separable within the hippocampal formation, the clusters we identified may alternatively reflect the peaks of a gradient along the anterior-posterior hippocampal axis, with statistical regularities predominantly represented in anterior regions, and semantic information predominantly represented in posterior regions (see Strange et al. 2014 for a review). Due to spatial correlations inherent to fMRI data, it is not possible to completely disentangle a gradient from two separable clusters.”(in Discussion, page 19).

Last, following the reviewer’s suggestion, we have added subtitles to Figures 2 and 3 in the manuscript.

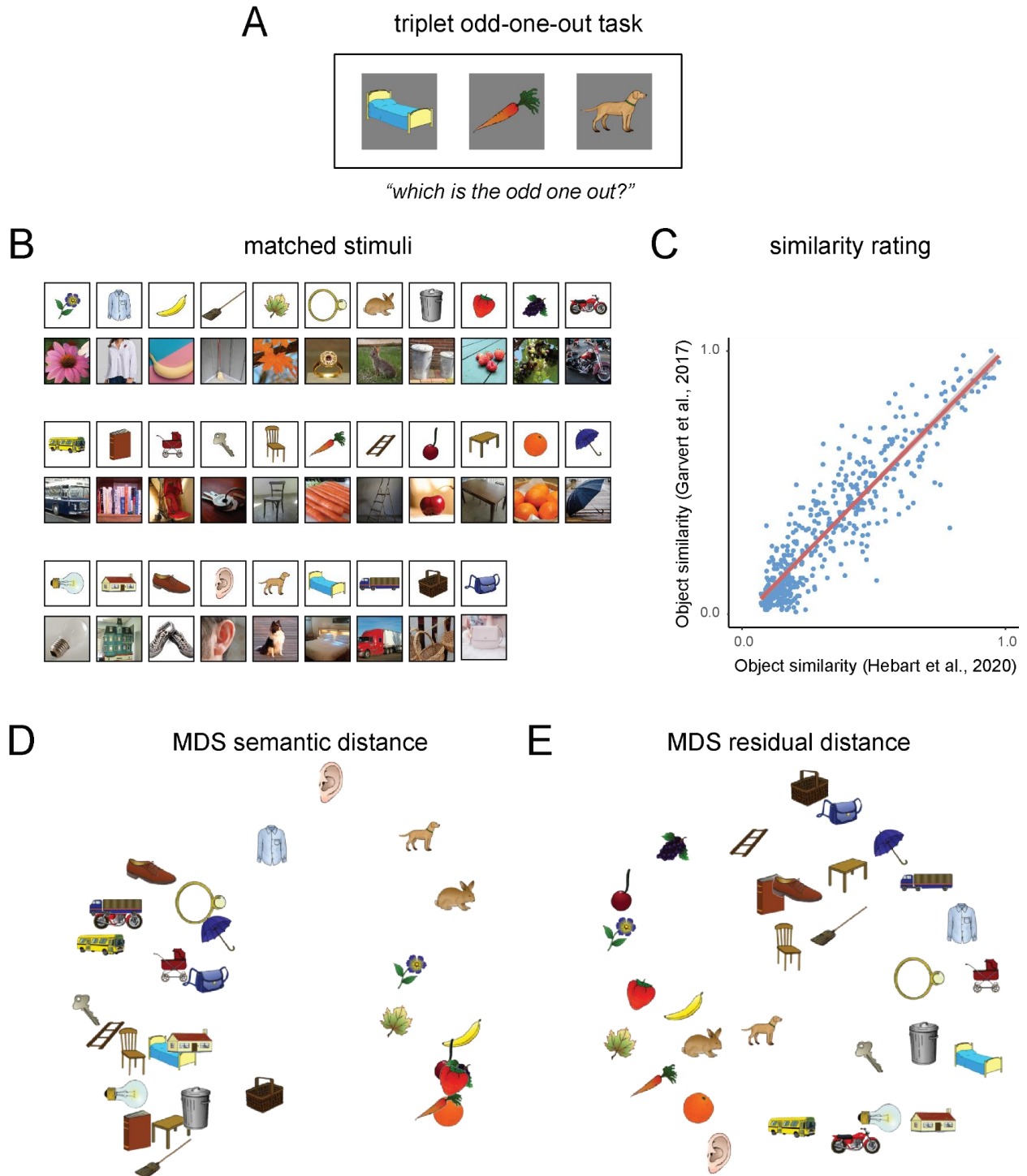


Figure 2. Semantic distance constructed using the triplet odd-one-out task. (A) An example trial of the triplet odd-one-out task. The task measures object similarity as the probability of participants choosing two objects together, irrespective of the context imposed by the third object (Hebart et al., 2020). (B) Stimuli used in the odd-one-out task. Top rows: all 31 stimuli from the

original study; bottom rows: a subset of stimuli from the THINGS database, matched with the 31 object stimuli used in the original study. The rating of the matched objects is done in the context of a total of 1,854 objects (Hebart et al., 2020). (C) Correlation between similarity ratings based on our own stimuli and ratings based on the corresponding stimuli from the THINGS database (Spearman's $Rho = .70, p < .001$). (D) Visualization of the 31 objects' semantic distance in a two-dimensional space according to multi-dimensional scaling (MDS). (E) 2D MDS visualization of the 31 objects' residual distance.

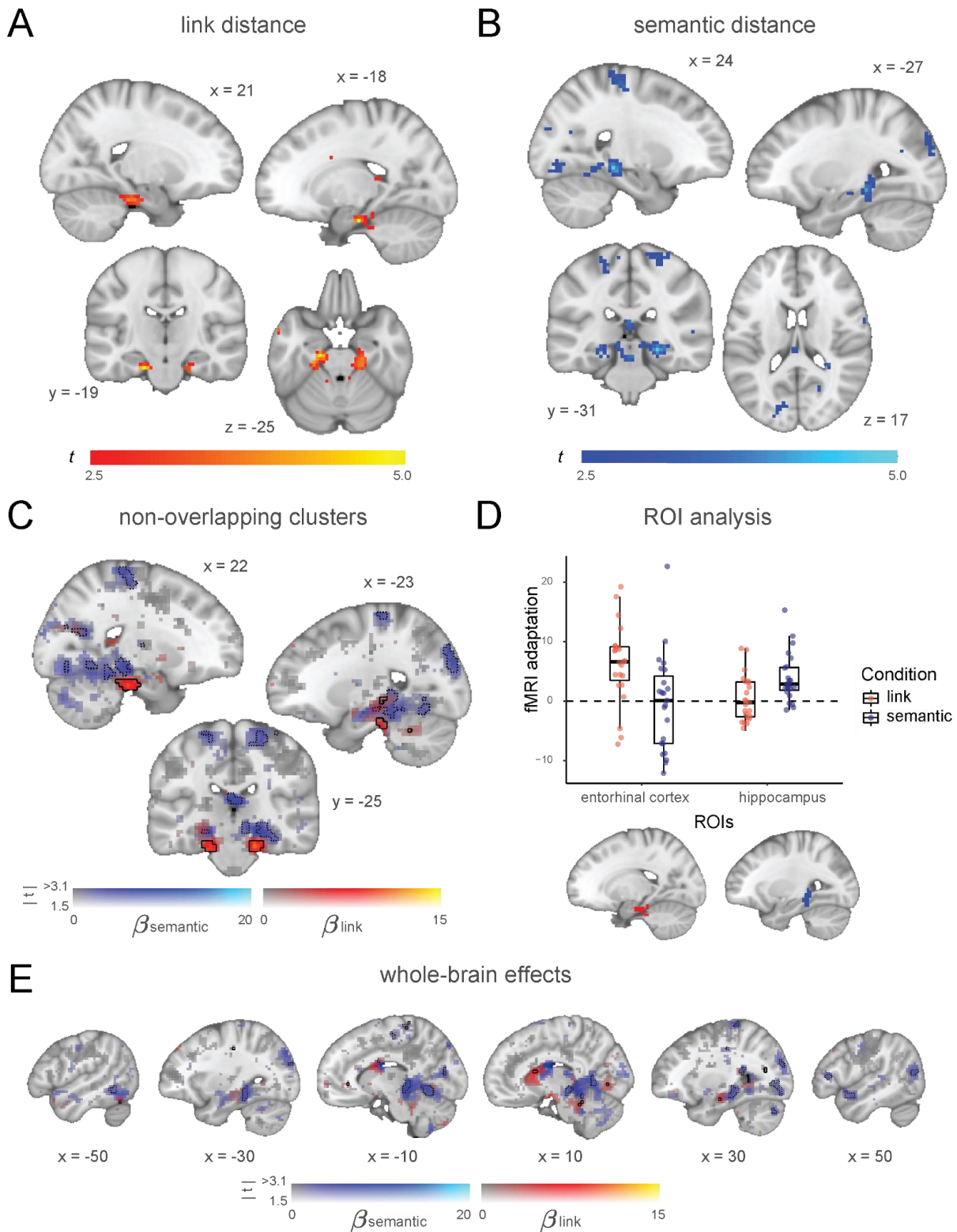


Figure 3. Statistical and semantic relationships are represented in non-overlapping clusters in the hippocampal-entorhinal system. (A) Whole-brain analysis showing a decrease in fMRI adaptation with link distance in the hippocampal formation, when link distance, semantic

distance and residual distance are included in the model. (B) Whole-brain analysis showing a decrease in fMRI adaptation with semantic distance in the hippocampal formation, when link distance, semantic distance and residual distance are included in the model. Both (A) and (B) are thresholded at $p < .01$, uncorrected for visualization. (C) Link distance effect (red) and semantic distance effect (blue) are represented in non-overlapping clusters. Whole-brain results are displayed using Slice Display (Zandbelt, 2017) which takes a dual-coding data visualization approach (Allen, Erhardt, & Calhoun, 2012), with color indicating the parameter estimates and opacity the associated t statistics. Solid and dotted contours outline voxels exceeding the cluster-defining threshold of $p < 0.01$ uncorrected for the link and the semantic effects, respectively. (D) Bottom: The two ROIs defined based on the link distance effect in the entorhinal cortex (in red) and the semantic distance effect in the hippocampus (in blue). Top: boxplot of the parameter estimates for the link distance and semantic distance effects extracted from the two ROIs. The thick horizontal line inside the box indicates the median, and the bottom and top of the box indicate the first and third quartiles of each condition. Each dot represents one participant. The plot is for visualization only, since the contrast used for defining the ROIs is not independent from the interaction effect of interest here. (E) Whole-brain results of link distance effect (red) and semantic distance effect (blue) are presented at equally spaced slices across the cortex, using Slice Display.

Responses to Reviewer #2

This manuscript explores the pressing question of how the brain represents different kinds of cognitive maps—specifically, whether the hippocampal formation integrates different dimensions into a single cognitive map, or whether different dimensions are captured in distinct, parallel maps. The present work explores these questions in the context of temporal statistical regularities vs. semantic similarities, and its most interesting finding is that these dimensions are captured in distinct maps localized to different regions in the hippocampal formation. The paper's main weakness is its failure to link these findings with our current understanding of semantic memory and to adequately consider the possibility that this distinction is due to different kinds of regularities (graph-based vs. Euclidean) rather than statistical vs. semantic content.

Reviewer #2 :

Zheng et al. reanalyze the data from Gavert et al. (2017) to ask how the hippocampal formation handles the embedding of stimuli with different relational structures. Specifically, they ask whether statistical regularities and semantic similarities are represented in one cognitive map or whether they are represented in distinct, parallel maps. They used an fMRI adaptation approach to replicate the finding from Gavert et al. that recently learned statistical regularities are represented in entorhinal cortex, and this manuscript's contribution is an analogous adaptation analysis to localize the representation of semantic similarities. They find that while statistical and semantic relations are both represented in the hippocampal formation, they are represented in distinct maps: statistical regularities are mapped in the entorhinal cortex, and semantic similarities are mapped in the hippocampus itself. The question of how different kinds of relations are represented in the brain as a whole—and medial temporal lobe specifically—is a pertinent one and this work is an important step to addressing that question. I appreciate that the authors modeled semantic similarities by finding the mutual information between participants' ratings of their own stimuli and ratings from a separate THINGS database which uses different naturalistic images of the same objects. I think this work is a valuable contribution, though I have some concerns with the framing and interpretation. These points are summarized below.

We thank the reviewer for their helpful feedback. We have addressed their comments below point by point.

1. The main results are framed as a distinction between statistical vs. semantic cognitive maps- the entorhinal cortex is claimed to represent statistics acquired through temporal transitions and the hippocampus is claimed to represent the semantic similarities between objects. However, can the distinct localizations of these "maps" clearly be interpreted in this way? There is another significant difference between the sought-for maps: statistical regularities were captured in a graph, with stimulus distance represented in terms of the number of links between the stimuli in the graph. On the other hand, semantic similarities were not captured in a graph-based format, but instead in a more Euclidean-based approach. This distinction between graph-based and Euclidean-based maps has been explored in other work (e.g., Peer et al., 2021, Structuring knowledge with cognitive maps and cognitive graphs, TiCS), and thus might be a relevant distinction that determines how information is structured within the hippocampal formation. What would the authors predict if statistical regularities were represented in a Euclidean map and semantic regularities were represented in a link-based graph? Would the localization of the representations swap, or does e.g., the hippocampus represent semantic information no matter if a graph- or Euclidean- approach is used? The authors briefly mention this possibility in a sentence at the end of the discussion, but this possibility could be taken more seriously given the authors' main claim (e.g. statistical vs. semantic distinctions).

We thank the reviewer for this insightful comment. It is an intriguing idea that the differences in anatomical localization we observed could be due to graph-like versus the Euclidean nature of the respective representations (see also Chown, Kaplan, & Kortenkamp, 1995; Kuipers, 2000; Mallot & Basten, 2009; Meilinger, 2008; Poucet, 1993 for similar discussions). After careful consideration, we have come to the conclusion that a difference in the nature of the type of knowledge we assess here is a plausible explanation for the anatomical separation we observe besides their age.

Perhaps this difference in the type of knowledge can be conceptualized in terms of the difference between conjunctive codes reflecting **objects embedded in a particular location or**

state - a coding pattern observed in the hippocampus -, as opposed to structural codes about the **relationship between states** represented in the entorhinal cortex (Whittington et al. 2021, Behrens et al. 2018).

In physical space, pattern similarity across voxels decreases with Euclidean distance. This is especially true in more anterior parts of the hippocampus where place cells are larger (Morgan et al., 2011). Hippocampal cells also encode abstract stimuli or concepts (Quiroga et al., 2005). Similar Euclidean distance effects are therefore to be expected for the semantic relationships between stimuli.

Entorhinal codes, on the other hand, likely reflect the transition structure between states rather than the states themselves. In fact, it has been suggested that entorhinal codes constitute a basis set for describing relational knowledge (Behrens et al., 2018). Such representations can best be described by graph-based distance measures such as communicability, a measure we found to reflect our data well in the original paper (Garvert et al., 2017).

If it was possible to represent statistical regularities in a Euclidean map and semantic regularities in a link-based graph, we would therefore indeed predict that the localization of the representations should switch, and the entorhinal cortex might represent semantic similarities between objects, while the hippocampus might represent statistical regularities based on the Euclidean map.

It is important to note though, that the communicability effect reported in Garvert et al. (2017) extended into the hippocampus, and lowering the threshold on the link distance measure also revealed hippocampal clusters (Figure R2-1). While we do now speculate about the reason for the predominant localization of statistical knowledge in the entorhinal cortex and semantic knowledge in the hippocampus, we also toned down strong statements regarding the precise special localization of the semantic and statistical knowledge within the hippocampal formation throughout the paper.

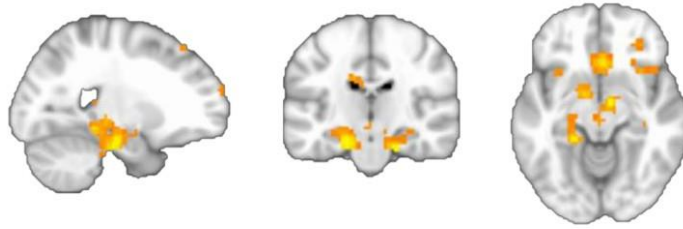


Figure R2-1. Whole-brain analysis showing a decrease in functional magnetic resonance imaging adaptation with link distance in the hippocampal–entorhinal system, thresholded at $p < 0.05$, uncorrected for visualization. Image from Garvert et al., 2017, author response image2.

We added the following text to the discussion:

“Alternatively, the segregation of the two maps may reflect differences in the learning process by which statistical versus semantic relational knowledge is acquired, or the nature of the underlying knowledge structures (Peer et al. 2021). It has been suggested that hippocampal cells encode conjunctive representations of objects in space, such that they are only active when a specific object is encountered in a particular location (Behrens et al. 2018; Komorowski et al. 2009; Wood et al. 1999). In the context of the current study, the population response of these cells may reflect objects embedded in semantic structures. More similar representations for objects suggest that they are closer to each other in a semantic space which reflects common associations or co-occurrences between objects. Entorhinal codes, on the other hand, likely reflect the transition structure between sequential states rather than the states themselves (Whittington et al., 2022). In our study, the relationships between objects can be characterized in terms of a graph structure reflecting transition probabilities between objects. Entorhinal grid cells may represent a low-dimensional basis set of a predictive representation of likely future locations in this state space (Stachenfeld et al., 2014; Stachenfeld et al., 2018). Indeed, the distance metric that best explained BOLD responses to be communicability in Garvert et al. (2017) was a graph-theoretic measure capturing the distribution of future states in a graph. It is closely related to the successor representation, a predictive encoding scheme of states in a reinforcement learning world (Dayan, 1993).

Our findings are thus consistent with the idea that the entorhinal cortex encodes statistical information about relational structure, while the hippocampus represents semantic similarities between specific objects.” (pages 18-19)

We now also adjusted the title to reflect this change. The article is now called:

Parallel cognitive maps for statistical and semantic relationships in the hippocampal formation

References:

Peer, M., Brunec, I. K., Newcombe, N. S., & Epstein, R. A. (2021). Structuring knowledge with cognitive maps and cognitive graphs. *Trends in Cognitive Sciences*, 25(1), 37-54. <https://doi.org/10.1016/j.tics.2020.10.004>

Wood, E. R., Dudchenko, P. A., & Eichenbaum, H. (1999). The global record of memory in hippocampal neuronal activity. *Nature*, 397(6720), 613-616. <https://doi.org/10.1038/17605>

Komorowski, R. W., Manns, J. R., & Eichenbaum, H. (2009). Robust conjunctive item–place coding by hippocampal neurons parallels learning what happens where. *Journal of Neuroscience*, 29(31), 9918-9929. <https://doi.org/10.1523/JNEUROSCI.1378-09.2009>

Morgan, L. K., MacEvoy, S. P., Aguirre, G. K., & Epstein, R. A. (2011). Distances between real-world locations are represented in the human hippocampus. *Journal of Neuroscience*, 31(4), 1238-1245. <https://doi.org/10.1523/JNEUROSCI.4667-10.2011>

Behrens, T. E. J., Muller, T. H., Whittington, J. C. R., Mark, S., Baram, A. B., Stachenfeld, K. L., & Kurth-Nelson, Z. (2018). What is a cognitive map? Organizing knowledge for flexible behavior. *Neuron*, 100, 490–509. <https://doi.org/10.1016/j.neuron.2018.10.002>

Stachenfeld, K. L., Botvinick, M. M., & Gershman, S. J. (2017). The hippocampus as a predictive map. *Nature Neuroscience*, 20(11), 1643-1653. <https://doi.org/10.1038/nn.4650>

Dayan, P. (1993). Improving generalization for temporal difference learning: The successor representation. *Neural Computation*, 5(4), 613-624. <https://doi.org/10.1162/neco.1993.5.4.613>

2. Semantic similarities are known to be captured in cortical representations. Because the authors used an anatomical map restricted to the hippocampal formation, there is no way to know how the hippocampal semantic representations relate to semantic representations elsewhere in the brain. Are the semantic representations in hippocampus redundant with information elsewhere in the brain? Or is a distinct contribution of these hippocampal representations to semantic cognition? A full-brain analysis isn't necessarily required, but an acknowledgement of semantic representations in cortex and the work that has been done in this field seems appropriate.

We set up the current study to address the key question of whether the hippocampal-entorhinal system which codes for statistical regularity also codes for long-term semantic knowledge.

Therefore, the focus has been put on the semantic representation within the hippocampal formation.

However, we also provided a full-brain analysis in the original manuscript, which would have identified areas outside the hippocampal formation if these also represented semantic distance.

“We report our results at a cluster-defining statistical threshold of $p < .01$ uncorrected, combined with small-volume correction (SVC) for multiple comparisons (peak-level family-wise error (FWE) corrected at $p < .05$). Activations in other brain regions were only considered significant at a level of $p < .001$ uncorrected if they survived whole-brain FWE correction at the cluster level ($p < .05$).” (page 9).

Our results showed no cortical regions whose adaptation signals change as a function of semantic distance after correcting for multiple comparisons on the whole-brain level. We additionally plotted the semantic distance effect (as well as the link distance effect) with equal-spaced slices across the brain (Figure R2-2), aiming to provide more information regarding regions outside the hippocampal formation. This indicates that the reported semantic effect is specific to the hippocampal formation. We discuss it in relation to the task relevance (see below). This plot has now been added to Figure 3 in the manuscript.

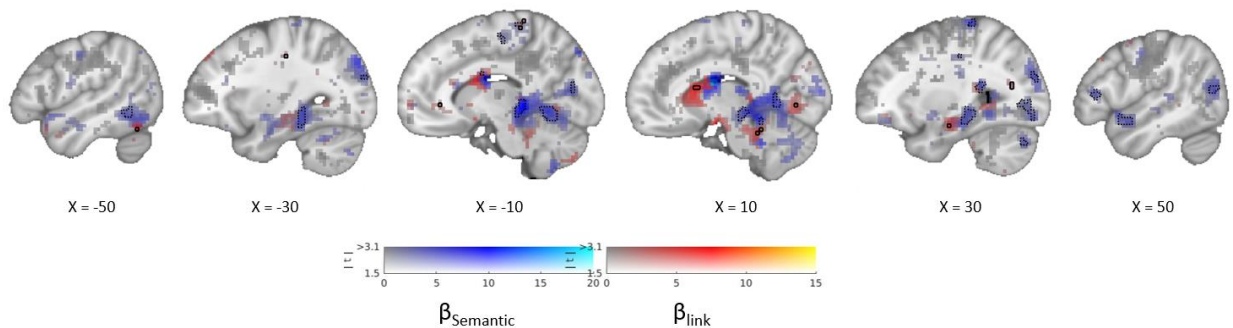


Figure R2-2. Link distance effect (red) and semantic distance effect (blue) presented at equally spaced slices across the cortex. Whole-brain results are displayed using Slice Display (Zandbelt, 2017) which takes a dual-coding data visualization approach (Allen, Erhardt, & Calhoun, 2012), with color indicating the parameter estimates and opacity the associated t statistics. Solid and

dotted contours outline voxels exceeding the threshold of $p < 0.01$ uncorrected for the link and the semantic effects, respectively.

We also added the following sentence to the Results section:

“No cortical regions showed fMRI adaptation effects as a function of semantic distance (all p s $> .71$, FWE corrected on the cluster level, Figure 3E).” (page 13).

We have additionally inspected the semantic distance effect specifically in two cortical areas which have previously been strongly associated with semantic processing: the left anterior temporal lobe (ATL, see Visser et al., 2010, *J. Cogn. Neurosci.* for a meta-analysis) and the left angular gyrus (see Humphreys et al., 2021, *TINS* for a recent review). ATL in particular is thought to play a domain-general role in retrieving semantic knowledge (Gorno-Tempini & Price, 2001; Schneider et al., 2018). Additional analysis with small volume correction using these anatomically defined masks yield no suprathreshold clusters for either the link distance or the semantic distance effect (Figure R2-3). This provides strong evidence that the reported semantic effect is specific to the hippocampal formation. In the discussion, we examined how this relates to task relevance (see below).

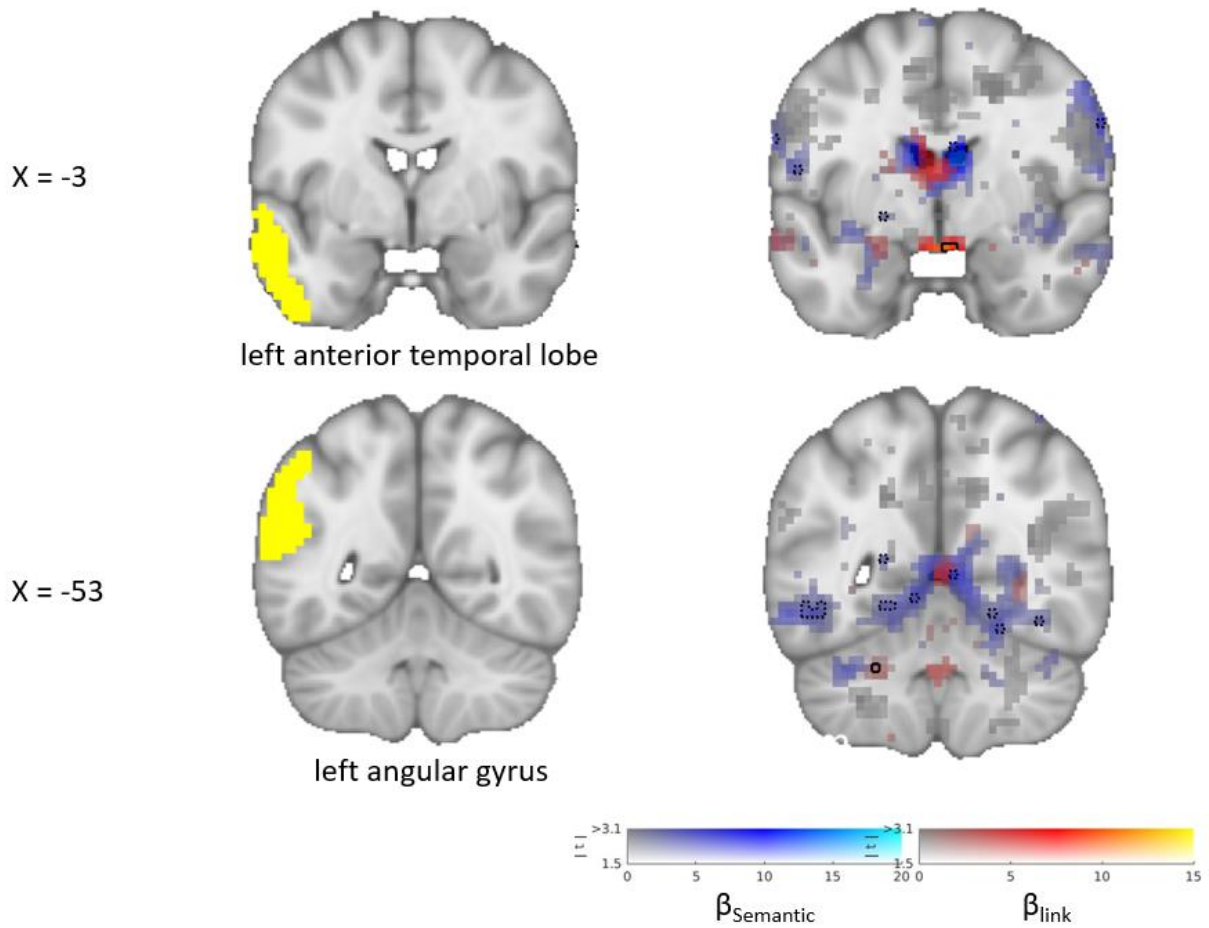


Figure R2-3. Two semantic-related ROIs used in the analysis (left panel) and corresponding data (right panel) representing link distance effect (red) and semantic distance effect (blue). Whole-brain results are displayed using Slice Display (Zandbelt, 2017) which takes a dual-coding data visualization approach (Allen, Erhardt, & Calhoun, 2012), with color indicating the parameter estimates and opacity the associated t statistics. Solid and dotted contours outline voxels exceeding the threshold of $p < 0.01$ uncorrected for the link and the semantic effects, respectively.

We also added this additional analysis to the manuscript.

“To explore the cortical semantic representation, we performed additional SVC using two anatomically defined masks: the left anterior temporal lobe and the left angular gyrus, two

regions previously reported to be important to semantic processing (Visser et al., 2010; Humphreys et al., 2021).” (in Methods, pages 9-10)

“No cortical regions showed fMRI adaptation effects as a function of semantic distance (all $ps > .71$, FWE corrected on the cluster level, Figure 3E), including specific regions of interest which have previously been associated with semantic processing: the left anterior temporal lobe (ATL, see Visser et al., 2010, for a meta-analysis) and the left angular gyrus (see Humphreys et al., 2021, for a recent review). ATL in particular is thought to play a domain-general role in retrieving semantic knowledge (Gorno-Tempini & Price, 2001; Schneider et al., 2018).

Nonetheless, additional analyses with SVC using these anatomically defined masks yielded no suprathreshold clusters for either the link distance or the semantic distance effect, indicating that the reported semantic effect is specific to the hippocampal formation.” (in Results, pages 13-14)

References:

Visser, M., Jefferies, E., & Lambon Ralph, M. A. (2010). Semantic processing in the anterior temporal lobes: a meta-analysis of the functional neuroimaging literature. *Journal of Cognitive Neuroscience*, 22(6), 1083-1094. <https://doi.org/10.1162/jocn.2009.21309>

Humphreys, G. F., Lambon Ralph, M. A., & Simons, J. S. (2021). A unifying account of angular gyrus contributions to episodic and semantic cognition. *Trends in Neurosciences*, 44(6), 452-463. <https://doi.org/10.1016/j.tins.2021.01.006>

Gorno-Tempini, M. L., & Price, C. J. (2001). Identification of famous faces and buildings: a functional neuroimaging study of semantically unique items. *Brain*, 124(10), 2087-2097. <https://doi.org/10.1093/brain/124.10.2087>

Schneider, B., Heskje, J., Bruss, J., Tranel, D., & Belfi, A. M. (2018). The left temporal pole is a convergence region mediating the relation between names and semantic knowledge for unique entities: Further evidence from a “recognition-from-name” study in neurological patients. *Cortex*, 109, 14-24. <https://doi.org/10.1016/j.cortex.2018.08.026>

We also agree with the reviewer that it is important to link our findings to the large body of literature on semantic representation. We have now added more detailed discussion embedding our findings into the previous work on semantic representation in the field:

“Our finding that a semantic map is represented in the hippocampus is also consistent with previous findings that the hippocampus is involved in the retrieval of semantic memory, particularly for relational knowledge between concepts, and that hippocampal activity reflects distances in semantic spaces (Estefan et al., 2021; Romero, Barense, & Moscovitch, 2019;

Solomon et al., 2019). The hippocampus thus seems to support domain-general processing of semantic knowledge (Staresina et al., 2011; Ranganath & Ritchey, 2012; Morton et al. 2021). However, previous research on semantic representations often shows the additional involvement of a broader set of brain areas, including cortical regions such as the anterior temporal lobe, the angular gyrus, the inferior frontal gyrus and the fusiform gyrus (Bracci et al., 2015; Charest et al., 2014; Clarke & Tyler, 2014; Price et al., 2015; Huth et al., 2016; see Frisby et al., 2023 for a recent review). These regions are thought to be involved in various aspects of semantic processing, such as semantic categorization, semantic retrieval, and the integration of semantic and perceptual information (Binder et al., 2009; Bookheimer, 2002; Lambon Ralph et al., 2016; Visser et al., 2010). Importantly, none of these functions were task-relevant in our study. Participants were not even required to pay attention to the objects, as they only had to attend to the presence of a gray patch on the screen. It is likely that the other regions that are involved in processing of semantic knowledge only become involved in situations where semantic knowledge is more relevant to task performance (see also Martin et al., 2018).” (pages 16-17).

References:

- Estefan, D. P., Zucca, R., Arsiwalla, X., Principe, A., Zhang, H., Rocamora, R., Axmacher, N., & Verschure, P. F. M. J. (2021). Volitional learning promotes theta phase coding in the human hippocampus. *Proceedings of the National Academy of Sciences of the United States of America*, 118(10), 1–12. <https://doi.org/10.1073/pnas.2021238118>
- Solomon, E. A., Lega, B. C., Sperling, M. R., & Kahana, M. J. (2019). Hippocampal theta codes for distances in semantic and temporal spaces. *Proceedings of the National Academy of Sciences of the United States of America*, 116(48), 24343–24352. <https://doi.org/10.1073/pnas.1906729116>
- Morton, N. W., Zippi, E. L., Noh, S. M., & Preston, A. R. (2021). Semantic knowledge of famous people and places is represented in hippocampus and distinct cortical networks. *The Journal of Neuroscience*, 41(12), 2762-2779. <https://doi.org/10.1523/JNEUROSCI.2034-19.2021>
- Staresina, B. P., Duncan, K. D., & Davachi, L. (2011). Perirhinal and parahippocampal cortices differentially contribute to later recollection of object-and scene-related event details. *The Journal of Neuroscience*, 31(24), 8739-8747. <https://doi.org/10.1523/JNEUROSCI.4978-10.2011>
- Ranganath, C., & Ritchey, M. (2012). Two cortical systems for memory-guided behaviour. *Nature Reviews Neuroscience*, 13(10), 713-726. <https://doi.org/10.1038/nrn3338>
- Frisby, S. L., Halai, A. D., Cox, C. R., Lambon Ralph, M. A., & Rogers, T. T. (2023). Decoding semantic representations in mind and brain. *Trends in Cognitive Sciences*. <https://doi.org/10.1016/j.tics.2022.12.006>
- Bracci, S., Caramazza, A., & Peelen, M. V. (2015). Representational similarity of body parts in human occipitotemporal cortex. *The Journal of Neuroscience*, 35(38), 12977–12985. <https://doi.org/10.1523/JNEUROSCI.4698-14.2015>

Charest, I., Kievit, R. A., Schmitz, T. W., Deca, D., Kriegeskorte, N., & Ungerleider, L. G. (2014). Unique semantic space in the brain of each beholder predicts perceived similarity. *Proceedings of the National Academy of Sciences of the United States of America*, 111(40), 14565–14570. <https://doi.org/10.1073/pnas.1402594111>

Clarke, A., & Tyler, L. K. (2014). Object-specific semantic coding in human perirhinal cortex. *The Journal of Neuroscience*, 34(14), 4766–4775. <https://doi.org/10.1523/JNEUROSCI.2828-13.2014>

Huth, A. G., De Heer, W. A., Griffiths, T. L., Theunissen, F. E., & Gallant, J. L. (2016). Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature*, 532(7600), 453–458. <https://doi.org/10.1038/nature17637>

Lambon Ralph, M. A., Jefferies, E., Patterson, K., & Rogers, T. T. (2016). The neural and computational bases of semantic cognition. *Nature Reviews Neuroscience*, 18, 42–55. <https://doi.org/10.1038/nrn.2016.150>

Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex*, 19(12), 2767–2796. <https://doi.org/10.1093/cercor/bhp055>

Bookheimer, S. (2002). Functional MRI of language: new approaches to understanding the cortical organization of semantic processing. *Annual Review of Neuroscience*, 25(1), 151–188. <https://doi.org/10.1146/annurev.neuro.25.1.12701.142946>

Price, A. R., Bonner, M. F., Peelle, J. E., & Grossman, M. (2015). Converging evidence for the neuroanatomic basis of combinatorial semantics in the angular gyrus. *The Journal of Neuroscience*, 35(7), 3276–3284. <https://doi.org/10.1523/JNEUROSCI.3446-14.2015>

Martin, C. B., Douglas, D., Newsome, R. N., Man, L. L., & Barense, M. D. (2018). Integrative and distinctive coding of visual and conceptual object features in the ventral visual stream. *eLife*, 7, e31873. <https://doi.org/10.7554/eLife.31873>

3. Relatedly, when setting up their approach, the authors say: "In line with the decrease in fMRI adaptation as a function of link distance observed in the entorhinal cortex (Gavert et al., 2017), we reasoned that in areas representing object relationships (e.g., semantic relationships), fMRI adaptation should scale with the corresponding distance measures (i.e., semantic distance)." This wording seems to imply that the idea to use fMRI adaptation to index semantic distance is a novel one. However, this is simply not the case, as many other researchers have used fMRI adaptation methods to study semantic representations. Some examples include:

- Yee et al. (2010), fMRI-adaptation evidence of overlapping neural representations for objects related in function or manipulation. *NeuroImage*.
- Conca et al. (2021), In search of different categories of abstract concepts: a fMRI adaptation study. *Scientific Reports*.
- Kim et al. (2009), Adaptation to objects in the lateral occipital complex (LOC): Shape or

semantics? Vision Research.

- Bedny et al. (2008), Semantic Adaptation and Competition during Word Comprehension. Cerebral Cortex.

We apologize if this wording implied that the approach was novel, this is of course not the case. We have revised it accordingly. We now say:

“In line with the decrease in fMRI adaptation as a function of link distance observed in the hippocampal formation (Garvert et al., 2017) and previous work measuring semantic distance using fMRI (Bedny et al., 2008; Conca et al., 2021; Kim et al., 2009; Yee et al., 2010), we reasoned that in areas representing object relationships such as semantic relationships, fMRI adaptation should scale with the corresponding semantic distance measure.” (page 13).

4. An analysis that would add to the value of the manuscript is an exploration of potential gradients between statistical and semantic representations in the hippocampal formation. Is there any evidence of regions whose representations are influenced both by semantics and by temporal regularities? One can attempt to extract such patterns by examining Fig. 3C, but this can only go so far. For example, is there any superior-inferior gradient, or anterior-posterior gradient? Whether or not these analyses are run, the authors could spend some time discussing the implications of having two completely separable maps, or how they think these representations may interact with each other.

We thank the reviewer for their insightful suggestion. To explore any potential gradients between statistical and semantic representations in the hippocampal formation, we performed a principal component analysis (PCA) based on the group-level statistics. Specifically, we extracted the group-level t-statistics for both the statistical and the semantic effects using an anatomically defined ROI combining the hippocampal formation (incl. hippocampus, entorhinal cortex, subiculum). We vectorized the voxel-based data and performed PCA on the link distance effect and the semantic effect. The analysis yielded two principal components, with the first component explaining 73% of the variance and capturing a negative correlation between statistical and semantic representations – consistent with what we have reported in the manuscript. We projected the loadings of the first component back into the voxel space (Figure R2-4). As we can

see from Figure R2-4, there seems to be a clear posterior-superior to anterior-inferior gradient in the right hemisphere, which comes out a bit less strongly on the left side.

However, the figure can also be read in line with the alternative interpretation of two separate clusters within the hippocampal formation y – as argued in the original manuscript. There might be neurons that separately code for semantic or statistical information, and the linear change of loadings in the voxels might just be a result of averaging different types of neurons. Given the nature of the current fMRI data, it is hard to fully disentangle these two alternative hypotheses.

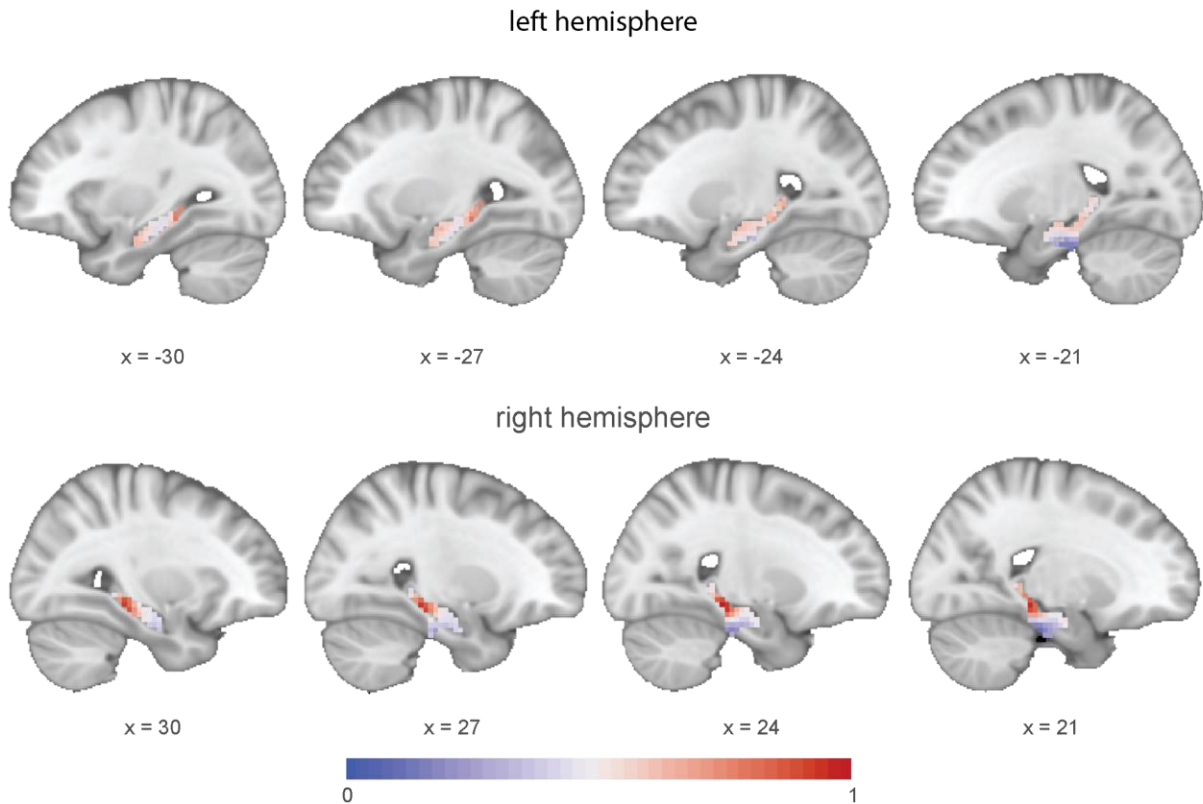


Figure R2-4. The first principal component from a principal component analysis of the statistical representation and the semantic presentation. The analysis is performed within the hippocampal formation (incl. hippocampus, entorhinal cortex, subiculum). A high loading means semantic representation is dominant, whereas a low loading means statistical representation is dominant. The loadings are normalized for visualization.

We now added this analysis to the manuscript. Figure R2-4 is included as a new Figure 4 in the manuscript.

“To further explore the spatial relationship between the link distance effect and the semantic distance effect, we sought to visualize their relationship in a hippocampal ROI that we defined anatomically (ie., same hippocampal mask used for SVC, combining hippocampus, entorhinal cortex, and subiculum). To this end, we conducted a principal component analysis (PCA) using group-level statistics for both the link and semantic distance effects. We extracted the group-level t-statistics for each effect using the anatomically defined hippocampal ROI. The voxel-based data was then vectorized and PCA was performed on the two distance effects. The resulting principal components were normalized and projected back into the voxel space for visualization.” (in Methods, page 10)

“To better understand the spatial relationship between the link distance effect and the semantic distance effect, we visualized their relationship within an anatomically defined hippocampal ROI. Specifically, we performed a PCA analysis of both effects within the ROI. This analysis yielded two principal components, with the first explaining 73% of the variance and capturing a negative correlation between statistical and semantic representations. As shown in Figure 4, there appears to be a posterior-superior to anterior-inferior gradient in the right hemisphere, is less pronounced in the left hemisphere.” (in Results, page 14)

We now discuss the possibility of a gradient as opposed to two separable clusters in the manuscript.

“Although these two maps seem to be anatomically separable within the hippocampal formation, the clusters we identified may alternatively reflect the peaks of a gradient along the anterior-posterior hippocampal axis, with statistical regularities predominantly represented in anterior regions, and semantic information predominantly represented in posterior regions (see Strange et al. 2014 for a review). Due to spatial correlations inherent to fMRI data, it is not possible to completely disentangle a gradient from two separable clusters.”(page 19).

In addition, as the reviewer suggested, we have included a more extensive discussion on the implications of having two separable maps.

“Cognitive maps have been proposed to be an organizing principle that underlies our ability to generalize and make inferences (Behrens et al., 2018)... However, while the two relational structures are represented in the same neural system, they are not represented in overlapping voxels. This suggests that the brain extracts separable relational structures in parallel rather than integrating them into one compositional map (Spiers, 2020). These parallel representations of separable maps likely facilitate generalization and inference in an ever-changing environment where the relevance of stimulus dimensions can shift rapidly. When different stimulus dimensions become relevant at different times, the parallel coding of multiple knowledge structures allows for flexible selection of relevant information. Such cognitive computations enable the hippocampus to adaptively generalize based on task demands (Garvert et al., 2023) and to guide goal-directed behavior in novel situations (Whittington et al., 2020). Additionally, attention can be selectively allocated to relevant state representations (Radulescu et al., 2021) and multiple relational structures can be flexibly combined into more complex compositional structures for generalization (Saanum et al., 2021).” (page 16)

References:

Behrens, T. E. J., Muller, T. H., Whittington, J. C. R., Mark, S., Baram, A. B., Stachenfeld, K. L., & Kurth-Nelson, Z. (2018). What is a cognitive map? Organizing knowledge for flexible behavior. *Neuron*, 100, 490–509. <https://doi.org/10.1016/j.neuron.2018.10.002>

Garvert, M. M., Saanum, T., Schulz, E., Schuck, N. W., & Doeller, C. F. (2021). Hippocampal spatio-temporal cognitive maps adaptively guide reward generalization. *bioRxiv*. <https://doi.org/10.1101/2021.10.22.465012>

Radulescu, A., Shin, Y. S., & Niv, Y. (2021). Human representation learning. *Annual Review of Neuroscience*, 44, 253-273. <https://doi.org/10.1146/annurev-neuro-092920-120559>

Saanum, T., Schulz, E., & Speekenbrink, M. (2021). Compositional generalization in multi-armed bandits. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 43, No. 43). <https://escholarship.org/uc/item/5nn9q6zc>

Spiers, H. J. (2020). The Hippocampal Cognitive Map: One Space or Many? *Trends in Cognitive Sciences*, 24(3), 168–170. <https://doi.org/10.1016/j.tics.2019.12.013>

Whittington, J. C. R., Muller, T. H., Mark, S., Chen, G., Barry, C., Burgess, N., & Behrens, T. E. J. (2020). The Tolman-Eichenbaum Machine: Unifying Space and Relational Memory through Generalization in the Hippocampal Formation. *Cell*, 183(5), 1249–1263.
<https://doi.org/10.1016/j.cell.2020.10.024>

Strange, B. A., Witter, M. P., Lein, E. S., & Moser, E. I. (2014). Functional organization of the hippocampal longitudinal axis. *Nature Reviews Neuroscience*, 15(10), 655-669.
<https://doi.org/10.1038/nrn3785>

5. The authors suggest that statistical regularities are represented separately from semantic regularities because of the "age" of the acquired regularities. Entorhinal cortex represents new statistical regularities, whereas hippocampus represents semantic regularities acquired across a lifetime. How can this view be reconciled with findings that statistical regularities are rapidly acquired in the hippocampus, and specifically hippocampal CA1 ? For example:

- Schapiro et al. (2016) Statistical learning of temporal community structure in the hippocampus. *Hippocampus*.

- Schapiro et al. (2017), Complementary learning systems within the hippocampus: a neural network modeling approach to reconciling episodic memory with statistical learning. *Philosophical Transactions of the Royal Society B*.

- Henin et al. (2021), Learning hierarchical sequence representations across human cortex and hippocampus. *Science Advances*.

We would like to thank the reviewer for this comment. Indeed, as the reviewer pointed out, there is evidence that statistical regularities can be rapidly acquired in the hippocampus. For example, the studies by Schapiro et al. (2016, 2017) and Henin et al. (2021) show that the hippocampus can learn the temporal structure of a sequence of auditory or visual stimuli and even temporal community structures over short timescales such as one experimental session. In contrast to this, in our original paper, we found the representation of the recently acquired map-like structure to be predominantly located in the entorhinal cortex (Garvert et al., 2017), and the representation of older semantic knowledge to be restricted to the hippocampus.

This point, together with the reviewer's comment about the nature of the representations made us reconsider our original argument that this anatomical separability is due to the age of the acquired regularities. Instead, we agree with the reviewer that a plausible explanation for the

anatomical separability is the different nature of the relevant relational structures and their mode of acquisition, with one reflecting transition probabilities and the other one reflecting concept similarities (see our response to your comment 1 above).

This still leaves us with the conundrum of why we did not observe any representation of statistical regularities in the hippocampus itself given that this has been observed in previous studies - an issue that was also raised in response to the original publication. We agree with the reviewers that this would have been a reasonable expectation. And indeed, if we lower the statistical threshold to (an overly liberal) value of $p < 0.05$, then we do see effects in the hippocampal formation as well, so we refrain from making strong statements about the precise localisation within the hippocampal formation.

Nevertheless, we would like to speculate about the reason why the hippocampal effects could be less pronounced here than in previous studies. One possible explanation for this is the fact that we explicitly looked for the representation of complex relational structures that require the integration over multiple pairwise relationships. The hippocampus encodes simpler statistical regularities such as transition probabilities between two stimuli (Shapiro et al., 2012) and even event boundaries such as those experienced during statistical sequences that follow a community structure (Shapiro et al., 2016), but it does not necessarily represent general geometric rules about state transitions by integrating multiple pairwise associations into more complex map-like structures.

We also agree with the reviewer that the timescale of acquisition is unlikely to be the only explanation for this segregation of statistical vs. semantic maps. Following the reviewer's suggestions (see also comment 1), we have now added more detailed discussion on alternative explanations (e.g., the different nature of the relevant relational structures, see our response to your comment 1 above), and on why the effect of statistical regularities is less pronounced in the hippocampus compared to previous studies.

“[One potential reason for this segregation may relate to a difference in the “age” of the acquired regularities. Whereas the statistical regularities were learnt on the day prior to scanning, the semantic relationships were acquired over the course of one’s lifetime.] This is, however, difficult to reconcile with the observation that statistical regularities, such as temporal structure of a sequence of auditory or visual stimuli and even temporal community structures can be

rapidly acquired in the hippocampus (Schapiro et al., 2016, 2017; Henin et al., 2021). One possible explanation for such a difference is the fact that we explicitly looked for the representation of complex relational structures that require the integration over multiple pairwise relationships. The hippocampus encodes simpler statistical regularities such as transition probabilities between two stimuli (Shapiro et al., 2012) and even event boundaries such as those experienced during statistical sequences that follow a community structure (Shapiro et al., 2016), but it might not necessarily represent general geometric rules about state transitions by integrating multiple pairwise associations into more complex map-like structures.” (pages 17-18)

References:

Schapiro, A. C., Turk-Browne, N. B., Norman, K. A., & Botvinick, M. M. (2016). Statistical learning of temporal community structure in the hippocampus. *Hippocampus*, 26(1), 3-8. <https://doi.org/10.1002/hipo.22523>

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

Henin, S., Turk-Browne, N. B., Friedman, D., Liu, A., Dugan, P., Flinker, A., ... & Melloni, L. (2021). Learning hierarchical sequence representations across human cortex and hippocampus. *Science advances*, 7(8), eabc4530. <https://doi.org/10.1126/sciadv.abc4530>

In general, the authors' findings interesting and valuable. However, I find their interpretations of their findings to be brief and shallow in that they do not relate their findings to the larger literatures of how the brain represents semantic information, how the hippocampus deals with graph-based vs. Euclidean similarities, and how statistical regularities may be rapidly acquired in the hippocampus.

We would like to thank the reviewer again for their helpful comments and suggestions. We hope the revised manuscript has addressed the points mentioned above and made a better link to literature in the field.