

Reviewed Preprint

v1 • July 30, 2026

Not revised

✉ For correspondence:

anna.shpektor@gmail.com

* These authors contributed equally to this work

Competing interests: No competing interests declared

Funding: See [page 29](#)

Reviewing editor: Anna C Schapiro, University of Pennsylvania, United States

© 2026, Shpektor et al. This article is distributed under the terms of the

[Creative Commons Attribution](#)

[License](#), which permits unrestricted use and redistribution provided that the original author and source are credited.

A hierarchical coordinate system for sequence memory in human entorhinal cortex

Anna Shpektor^{1,*}✉, Jacob JW Bakermans^{2,*}, Alon B Baram¹, Johannes Sarnthein^{3,4}, Debora Ledergerber^{4,5}, Lukas Imbach^{4,5}, Alina Kiseleva⁵, Emma Müller-Seydlitz¹, Helen C Barron^{1,6}, Timothy EJ Behrens^{1,2,7}

¹Wellcome Centre for Integrative Neuroimaging, University of Oxford, Oxford, United Kingdom • ²Sainsbury Wellcome Centre for Neural Circuits and Behaviour, University College London, London, United Kingdom • ³Klinik für Neurochirurgie, Universitätsspital Zürich, Universität Zürich, Zurich, Switzerland • ⁴Zurich Neuroscience Center (ZNZ), University of Zurich and ETH Zurich, Zurich, Switzerland • ⁵Swiss Epilepsy Center, Klinik Lengg, Zurich, Switzerland • ⁶Medical Research Council Brain Network Dynamics Unit, Nuffield Department for Clinical Neurosciences, University of Oxford, Oxford, United Kingdom • ⁷Wellcome Centre for Human Neuroimaging, University College London, London, United Kingdom

eLife Assessment

This **valuable** study provides fMRI evidence for a hierarchical representation of sequence, arranged along the anterior-posterior axis of the entorhinal cortex, as well as evidence from human single-unit recordings for sequence position coding in the hippocampus and entorhinal cortex. Weaknesses include tenuous links between the single-unit and fMRI datasets, **incomplete** evidence for some claims, lack of discussion of relevant hippocampal literature, and potential over-interpretation of the findings in relation to rodent grid coding. This work will be of interest to researchers studying memory, sequence representations, and spatial coding in the medial temporal lobe.

<https://doi.org/10.7554/eLife.112067.1.sa3>

Abstract

The entorhinal cortex (EC) supports a coordinate system for spatial memories, organised in a hierarchy along the EC dorso-ventral axis. Recent theories suggest that a similar coordinate system could scaffold non-spatial memories. Here we show that an abstract hierarchical coordinate system supports arbitrary sequence memories in the human medial temporal lobe (MTL). In single-unit recordings from MTL, we find abstract, coordinate-like coding in a simple sequential memory task. In fMRI we find that abstract coordinate representations are arranged hierarchically along the entorhinal cortex, mirroring the anatomical gradient of grid cells in the rodent EC but now for non-spatial sequences. We replicate this finding in an independent cohort of participants. These data suggest that memories are scaffolded on a hierarchical coordinate system aligned to preserved anatomy across domains and species.

Introduction

In fields like mathematics, physics or geography, a coordinate system is a tool constructed to determine the position of an object. To be useful, a coordinate system should be 1) efficient: compressing the position information effectively; and 2) generalisable: appropriate in different situations. Research in the well-studied field of spatial neuroscience has revealed that the brain has constructed a neural coordinate system for knowing one's position in physical space: grid cells in the entorhinal cortex¹, which are active on a regular triangular lattice when an animal moves in 2D space. Grid cells are arranged into discrete, factorised modules of different scales and spatial

frequencies². Readout from these modules can be combined to efficiently encode current location³ and to support inferences such as shortest path to goal⁴. The grid cells code also generalises: it abstracts over the sensory particularities of spatial environments^{5–7}. Beyond space, neurons in the entorhinal cortex have also been shown to provide a temporal basis for ongoing experience^{8–10}, though whether such codes are organised along an anatomical hierarchy is not known.

Here we set out to ask whether the same coding principles extend beyond physical space to organise arbitrary sequences of experience. Such a coordinate system could serve as a scaffold¹¹ for non-spatial memories. For example, the position in the sequence breakfast-lunch-dinner can be thought of as a “stage of day” coordinate and provides a useful reference to encode and retrieve memories. Encoding this sequence with a coordinate entails a generalisable code: all breakfasts share commonalities, enabling abstraction that can aid behaviour in novel situations. Importantly, sequences often have a nested hierarchical structure, with different levels of the hierarchy corresponding to different scales or temporal frequencies. This can allow for an efficient factorial description in a coordinate system with multiple coordinates, akin to grid modules in physical space or to the multiple timescales encoded by entorhinal neurons in time. For example, consider the hierarchy “day stages-days-weeks-months-years”, where each level consists of a sequence from the previous level. You might remember that the first astronauts stepped foot on the moon in 1969, but not that they did so in July. Conversely, you might look back on your vacation in Italy and be confident that you travelled in August, but not remember in which year. A representation of such a hierarchy with a coordinate system can use the structure in sequences to encode them usefully for memory.

We first tested for a representation of a single coordinate of task stage by recording from neurons in the MTL of patients performing a simple sequential memory task. We find cells responding to different task stages, similar to previously observed time cells, but abstracted over sensory input. We then use fMRI in two independent cohorts of healthy volunteers to probe a factorised, multiple coordinate representation of general hierarchical sequences, serving as a scaffold for memory. We aimed to test whether, like the entorhinal codes for space and time, sequence representations are abstract, and organised on an anatomical hierarchy. While our hypotheses are inspired by the properties of spatial grid cells and lateral entorhinal temporal codes, we note that our study did not address whether the proposed sequence coordinate system repurposes these cells or involves different ones. Moreover, while grid cells are organised into discrete modules, our task could in principle be represented either as discrete modules or as a continuous gradient.

Results

Representation of task stage coordinate in human MTL

Knowing one’s position in the task at hand is particularly useful in tasks with a repeating structure – as is common in lab tasks with same-structured trials. To test if the human MTL encodes a “task stage coordinate” representation in such a task, we recorded cells in hippocampus (92 neurons) and entorhinal cortex (75 neurons) of epilepsy patients performing a memory task with a repeating sequential structure (Fig S1.1 [↗](#) and Fig S1.2 [↗](#)). In each trial, patients were shown four pictures in a sequence, and after a delay of 3 seconds were asked to report the sequence position of one of the pictures (Fig 1a [↗](#)). Importantly, while the same 4 pictures were used in every trial, their order was randomised across trials (Fig 1b [↗](#)). Patients performed the task well, with an average accuracy of $93.7\% \pm 6.3\%$ (SD).

Visualisation of neural population trajectories strongly suggested that a task stage coordinate was represented, as the activity for different task stages (initiation, stim1, stim2, stim3, stim4, delay, question, feedback) separated in neural space (Fig 1c [↗](#)). Indeed, using a Bayesian decoder, we could decode each task stage from the population activity with high accuracy (Fig 1d [↗](#), average decoding accuracy of correct task stage 0.73, chance is 0.125). This representation was also present at the single neuron level: A one-way ANOVA revealed that most cells (Hippocampus: 77%, EC: 88%, Fig 1e [↗](#)) in the MTL were tuned to task stage (see example cells in Fig 1e,f [↗](#)). Together, these results indicate that neurons in the MTL overwhelmingly represent different stages of the

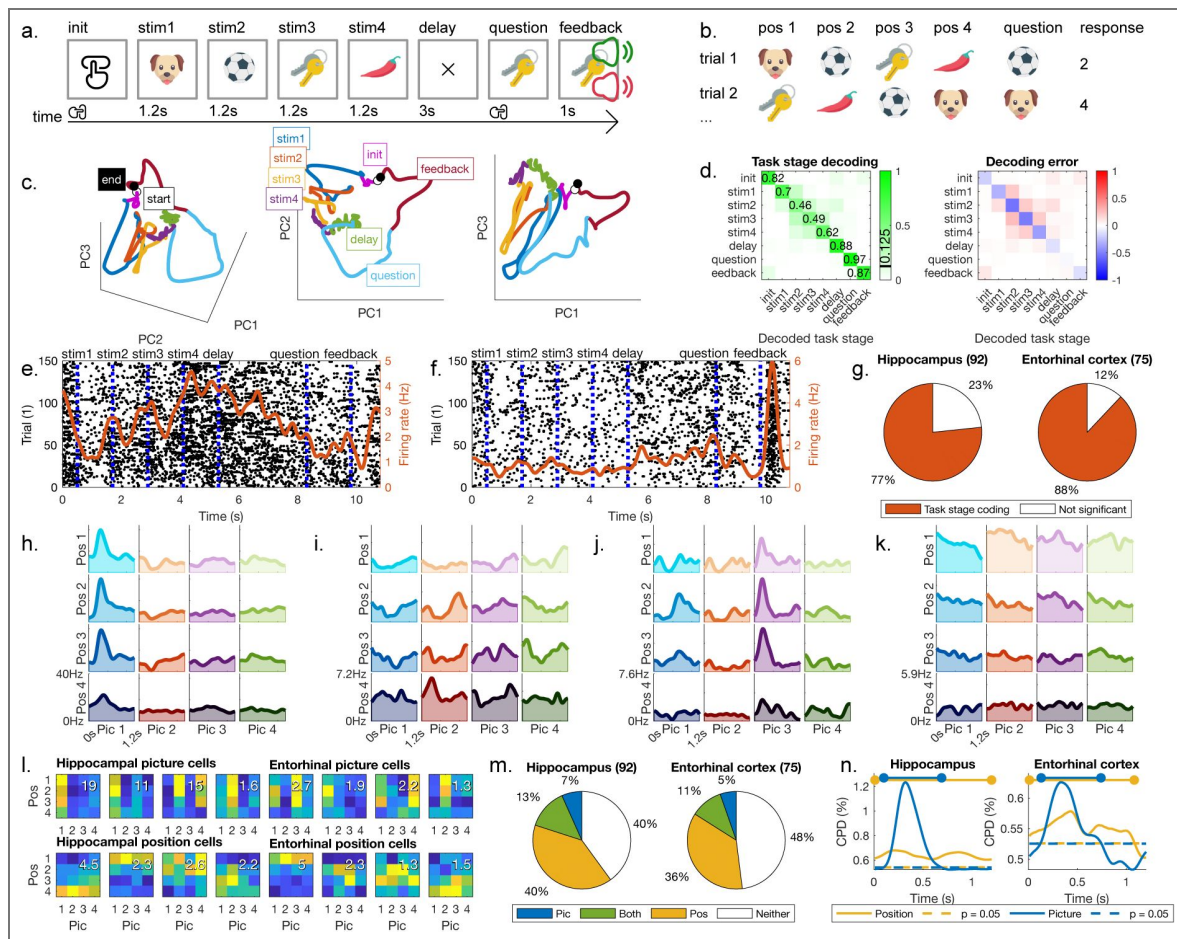


Fig 1. Single unit data.

a. Participants had to learn the order of 4 pictures presented one after another. After a 3-second delay, we queried the sequence position of one of the pictures. **b.** On each trial, we used the same four pictures but changed the order in which they were displayed. **c.** The averaged population trajectory throughout a full trial of the task in neural activity space shows that different task stages occupy different regions in neural activity space. **d.** Decoding task stage from the population firing rates. Left: average decoded task stage (columns) given true task stage firing rates (rows) across trials, with chance level (1/8) indicated on the color scale. Right: decoding error. **e.** Raster plot across trials for a hippocampal cell showing significant task-stage modulation. The red line is the average firing rate. **f.** Raster plot across trials for an entorhinal cell showing significant task-stage modulation. The red line is the average firing rate. **g.** Proportion of cells showing significant task-stage coding (initiation, stim1–4, delay, question, feedback; one-way ANOVA). **h–l.** Example picture or coordinate cells. We averaged the firing rate of each cell across all trials of a particular picture (column) in a particular position (row). **h.** Hippocampal cell coding for picture 1. **i.** Hippocampal coordinate cell that increases its firing from the first to the last sequence position. **j.** Entorhinal cell coding for picture 3. **k.** Entorhinal coordinate cell that decreases its firing from the first to the last sequence position. **l.** Other examples of picture (top) and position (bottom) specific cells in hippocampus (left) and entorhinal cortex (right) averaged across time within trial stage as a compact variant of the plots in h–k. The maximum average firing rate is indicated by the white number for each cell. **m.** Proportions of picture and position specific cells in hippocampus (left) and entorhinal cortex (right). **n.** Variance explained through time from stimulus presentation by picture and position factors on hippocampus (left) and entorhinal cortex (right).

task – across the population they can code the structure of the task. Note, however, that this representation is not necessarily abstract, as these analyses are partially confounded by external sensory experience (e.g. common visual experience in all trials during the initiation or delay periods, and common auditory experience during feedback) and internal neural processes (e.g. working memory, motivation, motor preparation).

Coordinate cells in the human MTL encode the position in a sequence abstractly

Next, we asked whether MTL coordinate cells abstracted over the sensory particularities of trials. To test this in a controlled manner, we focused on the presentation periods of the pictures during encoding (stim1-stim4), where our design ensured there were no visual and motor confounds. Such “abstract coordinate cells” would fire, for example, in the third position regardless of the picture shown. An ANOVA with two factors (position and picture) revealed a significant main effect of position in approximately half of the recorded MTL neurons (Fig 1m, Hippocampus: position only 40%, picture only 7%, both 13%; Entorhinal cortex: position only 36%, picture only 5%, both 11%; example cells in Fig 1g-l and Fig S1.3, S1.4; tuning to only position 2 and 3 in Fig S1.5). Interestingly, while the sensory picture code was (expectedly⁹) present transiently after picture presentation onset (blue line in Fig 1n, permutation $p < 0.05$ family-wise error corrected through time: hippocampus: 0.1-0.69s, EC: 0.13-0.74s), the abstract position code was present before and after the presentation of the stimulus (yellow line in Fig 1n, hippocampus: 0.01-1.2s, EC 0.01-1.08s). In summary, these results indicate that in this simple sequential memory task, the human MTL encodes an abstract coordinate representation, independent of sensory stimuli.

The hierarchical structure of sequences is used behaviourally as a scaffold for memory

Our next goal was to investigate whether the MTL's sequence coordinate system would extend to memory tasks with hierarchical sequential structures, where an efficient representation should comprise multiple coordinates – one for each level of the hierarchy. Indeed, the neural coordinate system for space (grid cells) is organised in exactly such a hierarchical form – grid cell modules (Fig 2a), which compress the spatial code in a factorised manner (Fig 2b). Critically, these modules are anatomically organised on a ventral-dorsal gradient in the rodent brain, from large scale (low frequency) to small scale (high frequency) grid modules² (Fig 2d top). Whether non-spatial sequences are organised along a similar gradient – by grid cells or other entorhinal cell types – is unknown. We hypothesised that hierarchical sequences used for memory will be represented in the human EC with a coordinate system organised on a similar anatomical gradient (Fig 2c) – whether or not this gradient relies on grid cells themselves. Such a finding would resonate with recent theoretical suggestions that grid cells are a compact abstraction of 2D sequences, and that similar principles could apply to sequences drawn from other structural forms^{7,12–15} (though potentially in different cells than grid cells).

Due to the tilt of the human MTL compared to the rodent's, the human EC gradient should extend along the anterior-posterior (A-P) rather than ventro-dorsal axis¹⁶ (Fig 2d bottom). Because obtaining enough cells along the A-P axis in human patients is impractical, we used fMRI. Although this study aimed to measure the organisation of memory, we decided not to rely on real remembered episodes of our participants, as these are easily distorted by e.g. emotional valence or recency biases. Instead, we designed a highly controlled task that induces a particular memory structure to afford precise representational hypotheses. We collected two independent datasets (26 and 22 participants, obtained years apart), and replicated all results across the two datasets.

We designed a task to look for hierarchical organisation of sequence coordinates using auditory sequences. This entailed teaching participants two 131-tones-long auditory sequences by repeated listening and testing of the sequences for 2 days (Fig 2e,f,g). The sequences shared the same nested hierarchical structure, with 4 levels – each level containing 3 or 4 locations (Fig 2f,g and Fig S2.1, Methods). Any specific (flat, one-dimensional) location in this sequence can therefore be pinpointed by 4 coordinates, where here a coordinate is a location within a level of the

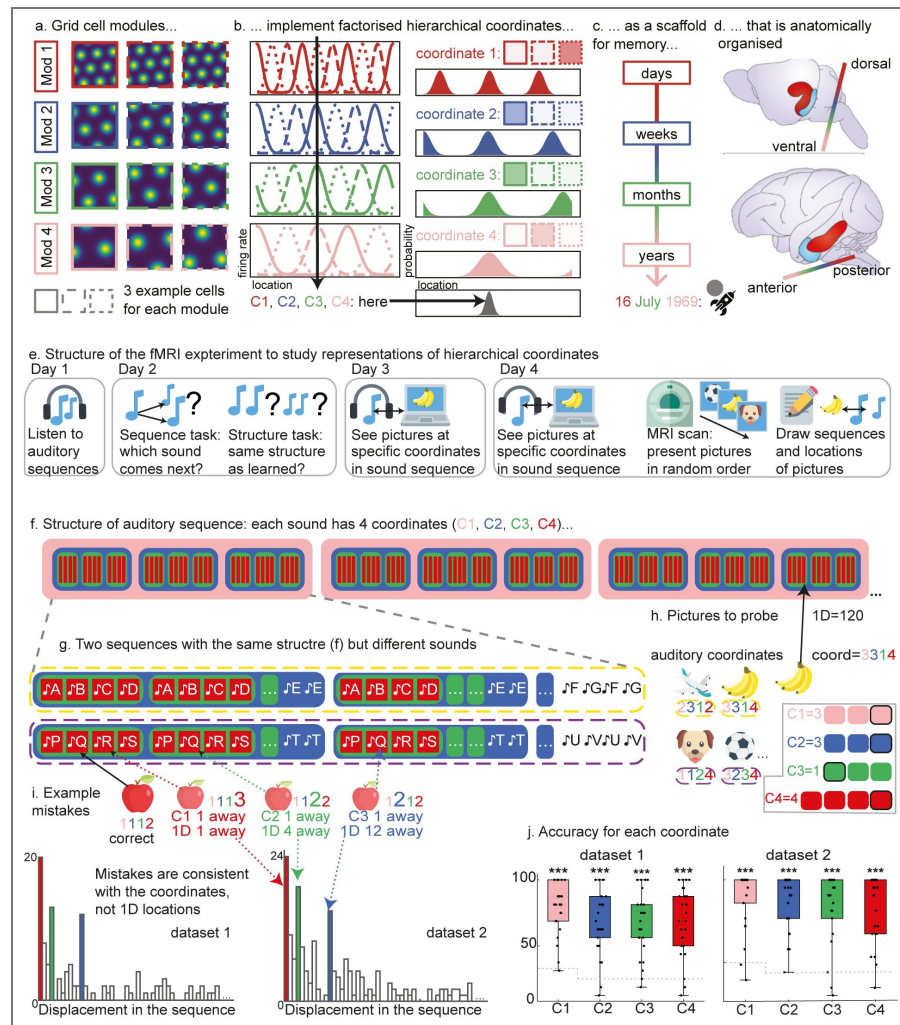


Fig 2. fMRI experimental design and behaviour.

a-d. Schematic of logic for anatomical hypothesis: coordinates systems organised anatomically by scale in EC. **a-b.** Grid cells are a coordinate system for space. **a.** Simulated example grid rate maps. Grid cells in rodent entorhinal cortex group together in anatomically separable modules of different spatial scales (rows). Cells within a module exhibit different shifts or 'phases' (columns, in differently dashed boxes). Adapted from Stemmler et al., 2015³. **b.** Together, these grid cell modules implement a factorised hierarchical coordinate system, here illustrated in 1D for clarity. Given the tuning curves of each grid cell in a module (left, different dashes), a particular firing pattern across cells (e.g. cell 3 in module 1 is active while cells 1 and 2 are silent) specifies a particular (module scale dependent) distribution over locations (right) - the readout of the coordinate. Combining the multi-scale coordinates across modules yields a hierarchical code for one specific location (grey, bottom). **c.** We hypothesise that similar factorised hierarchical codes in the human brain provide a scaffold for memories. For example, the hierarchical representation '16 July 1969' corresponds to the memory of the first man on the moon. **d.** For physical space, this gradient of coordinate representations from small to large scales is organised anatomically in rodent entorhinal cortex along the dorsal (small scale grid cells) to ventral (large scale grid cells) axis. In humans, this axis has rotated into the posterior-to-anterior direction. **e.** We test this hypothesis in a four-day neuroimaging experiment. Participants learn auditory sequences (day 1-2) with specific locations tagged by pictures (day 3). We show the same pictures in random order in the MRI scanner on day 4, without any auditory stimuli. After the scan, we ask participants to draw the sequences and place the pictures at their associated locations (an example drawing in Fig S2.2). **f.** Hierarchical structure of the auditory sequences (full sequences in Fig S2.1). **g.** Participants learned two sequences with identical structures but different sounds. **h.** Example pictures and their specific sequence locations that participants learned on the third day. The 1D position of every picture in the auditory sequence can be pinpointed using four coordinates. In the fMRI scanner we showed pictures in a random order without sounds, with the goal of reactivating the representations of the corresponding auditory coordinates. **i.** Distance between incorrectly placed pictures in the post-scan drawings of day 4 and their correct locations. The three strongest peaks in the histogram correspond to incorrectly identifying the coordinate in one level while correctly identifying all other coordinates. **j.** The average accuracy of the post-scan drawings across all pictures for each hierarchical coordinate.

hierarchy (Fig 2h [↗](#), right). We hypothesized that these coordinates would be represented along an anatomical gradient in the entorhinal cortex with coarsest level's coordinate anterior and finest level's coordinate posterior (Fig 2d [↗](#)).

However, to be certain that we could measure the coordinate representations without biases such as temporal autocorrelation, we needed to be able to present the coordinates in a random order. We therefore could not present the actual auditory sequence in the scanner. Instead, we tagged 8 locations in each sequence with visual images (Fig 2h [↗](#)). To do this, on the third day of auditory sequence learning, participants saw 8 images on a computer monitor at specific sequence locations during 25 repeats of the sequence (Fig 2e,h [↗](#)). This procedure therefore endowed 16 images (8-per-sequence) with 4 sequence coordinates each (one-per-level). On the fourth day, participants heard 10 more repeats of the sequence with the paired images as a reminder, and then underwent fMRI scanning where they viewed the images in a random order, without any auditory tones (Fig 2e [↗](#)). Crucially this meant any structure in the resulting fMRI activity could only be a result of the underlying coordinate representation, and not the statistics of the presentation order. To maintain attention, participants were instructed to respond when the shape of the image was distorted (6% distractor trials). Subsequent analyses look for the neural representations of the *images' coordinates* in entorhinal cortex.

After the scan, participants drew the sequences, placing the pictures in their correct positions (see Fig S2.2 [↗](#) for an example). Accuracy was significantly above chance for all levels (Fig 2j [↗](#), t-test against chance: all levels, both datasets $p < 0.001$). The displacement errors were consistent with a hierarchical rather than a flat representation of the sequence: the strongest peaks correspond to mistaking one coordinate of a picture while correctly placing all other coordinates (Fig 2i [↗](#)). Indeed, the distance in hierarchical coordinate space was a better predictor of errors across participants than linear sequence distance (multiple regression, $p < 0.001$ dataset 1, $p = 0.02$ dataset 2).

A coordinate system for hierarchical sequences is organised on an anterior-posterior gradient in the human entorhinal cortex

To test our hypothesis of an anterior-posterior EC gradient from low to high frequencies, we first needed to quantify the coordinate representation at each of the 4 levels around each EC voxel. Using representational similarity analysis^{17,18} (RSA), we built a hypothesis similarity matrix for each level, which assumed that nearby coordinates in that level would have more similar neural representations. This led to 4 hypothesis matrices (Fig 3b [↗](#)). For each level, the matrix was 16×16 , with each element being the predicted similarity between a pair of images' coordinates at that level. We used these as regressors in a multivariate RSA searchlight analysis¹⁹. The data matrix, at each voxel, was the correlation matrix between voxel responses in a local searchlight between the 16 images, resulting in a 16×16 data matrix per-voxel. We ran a multiple regression resulting in 4 t-statistic maps in each dataset – one per level (Fig 3c [↗](#), Methods).

No voxels survived whole brain comparison for any level (perhaps due to the passive nature of the task, or the extremely small areas of cortex hypothesized to respond to individual levels). Nevertheless, we were able to look for the relative positioning of the different levels' coordinate effects within entorhinal cortex, and to replicate these relative positions in an independent dataset. Indeed, visual inspection of both independent datasets suggested that for the left EC, these effects were ordered on the A-P axis from low to high frequencies (visualisation Fig S3.1a [↗](#)), as hypothesised (Fig 3c [↗](#): example participant, Fig 3e [↗](#): dataset 1, Fig 3f [↗](#): dataset 2, Fig 3g [↗](#): both datasets together).

To statistically test for the gradient, we first extracted the centre of mass (CoM) of each level's map (masked within each hemisphere's EC), and projected it on the anterior-posterior (A-P) axis, yielding a single A-P location per level in each participant (Y_i in Fig 3d [↗](#), left). We aimed to test whether the order of these projections matched the hypothesised order, in both datasets. We quantified order by calculating an "orderness" metric from distances between ordered pairs of A-P locations (Fig 3d [↗](#) right). This metric computes the sum of differences between the A-P locations

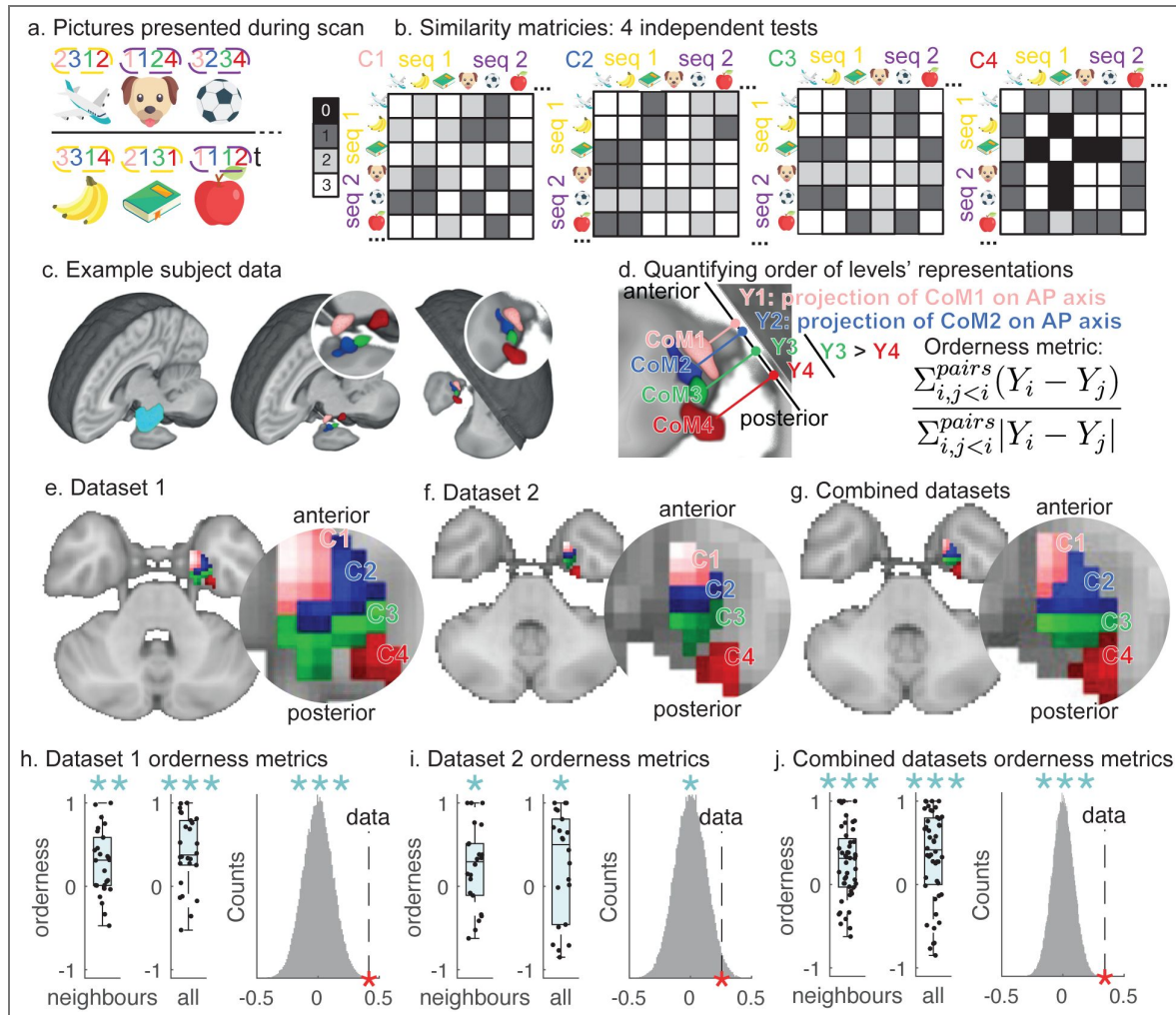


Fig 3. Coordinate representations are anatomically organised.

a. We present pictures associated to specific coordinates in random order in the MRI scanner. **b.** To find coordinate representations, we specify hypothesised similarities between pictures with respect to each level of the hierarchy. For example, the dog and the book have high similarity for coordinate C2 (blue: both 1), but low similarity for coordinate C4 (red: 4 versus 1). **c.** We use RSA to detect within entorhinal cortex (left) where neural similarities match the hypothesised similarities (example participant, middle), along the anterior-posterior axis (right). **d.** To quantify the A-P order of coordinate representations, we calculate an “orderness metric” from differences between projections (Y_i) of each coordinate representation’s centre of mass (CoMi) on the anterior-posterior axis. If the gradient is present, the differences are expected to be positive, because we always subtract the lower from the higher level, and the projection value increases in anterior direction. **e-g.** Visualisation of group averaged coordinate representations in dataset 1 (e), dataset 2 (f), and both combined (g). **h-j.** Orderness metric for neighbouring levels’ pairs only (left), including all pairs of levels (middle), and averaged across participants against a null distribution from shuffled level labels (right) in dataset 1 (h), dataset 2 (i), and both combined (j).

of pairs of levels' representations (positive if the low frequency level is anterior to the high frequency level, as hypothesised), normalised by the unsigned distance. This metric is more positive if low frequency levels are represented more anteriorly than high frequency levels. Indeed, the orderliness metric was positive for different combinations of sequence coordinates in both datasets, which confirmed our hypothesised order. First, we calculated the orderliness metric only for the neighbouring levels in the hierarchy (e.g. levels 1-2, 2-3 etc.). Indeed, they were ordered according to our hypothesis (Fig 3h-j left, one-sided t-test: dataset 1 $p < 0.001$, dataset 2 $p < 0.05$, combined $p < 0.001$, Supplementary Table 1). The orderliness metric including all pairs of levels was also positive (Fig 3h-j middle, one-sided t-test: dataset 1 $p < 0.001$, dataset 2 $p < 0.05$, combined $p < 0.001$, Supplementary Table 1). This measure was significantly larger than a null distribution created by randomly shuffling level labels (Fig 3h-j right, permutation test: dataset 1 $p < 0.001$, dataset 2 $p < 0.05$, combined $p < 0.001$, Supplementary Table 1). All results above are reported in the left entorhinal cortex; we did not find evidence of a gradient in right entorhinal cortex (Fig S3.2, Supplementary Table 2).

Beyond the replication across two independent datasets, the results were also not sensitive to the specific analysis choices we made. The results held when we used a more conservative EC mask (all $p < 0.05$, Fig S3.3, Supplementary Table 3), and when using a discrete measure of the A-P order of the levels rather than a continuous one (all permutations $p < 0.05$, Fig S3.4, Supplementary Table 4). Furthermore, the locations of the individual levels' coordinate representations replicated across datasets. On average, a level's t-statistics were higher when computed at that level's centre of mass from the other dataset than when computed at the centre of mass of other levels in the other dataset ($p < 0.01$, Fig S3.1b,c).

The EC's sequence coordinate system abstracts over sensory stimuli

So far, we have shown that the EC encodes a sequence coordinate system in an efficient hierarchical fashion, and that this coordinate system is organised anatomically in a similar way to grid cell modules - the EC's coordinate system for physical space.

Notably, the grid representation is also abstract – another important property of coordinate systems: within a grid module, the covariance structure of grid cells' activity is maintained across different spatial environments⁵. In other words, the information embedded in grid cells abstracts over sensory details. EC codes for time show a related form of abstraction, being reshaped by the structure of events rather than tracking sensory experience alone^{8,9}. To test whether this was also true for our sequence coordinates, we re-did the entire analysis above, but now selecting only the portions of the hypothesis matrices which compared images from *different auditory sequences* (Fig 4b). This means any similarities between the data and the coordinate hypothesis matrices could not be because of sequence-specific reasons. We then followed the same procedure for the analyses as described above (group statistical maps in Fig 3c-e). Despite the reduced power of the analysis (using only half of the hypothesized correlations), all permutation tests were significant (Fig 4f-h right, permutation test: dataset 1 $p < 0.001$, dataset 2 $p < 0.05$, combined $p < 0.0001$, Supplementary Table 6) indicating an abstract generalisable coordinate system.

Discussion

Understanding the temporal structure of experiences is a key aspect of learning and memory. Here, we show that the brain encodes this structure with an abstract coordinate system, anatomically organised hierarchically for efficient encoding. Using single-unit data, we first show that in a task with a repeating structure the MTL encodes a stimulus-invariant task-stage coordinate. We then show using fMRI that in a task with a hierarchical structure the EC encodes multiple factorised coordinates which are arranged anatomically on the anterior-posterior EC axis according to their level in the hierarchy, mirroring the organisation of grid cell modules in spatial tasks. Taken together, these results suggest that a coordinate system-like representation might be a general shared representational principle across multiple tasks and domains.

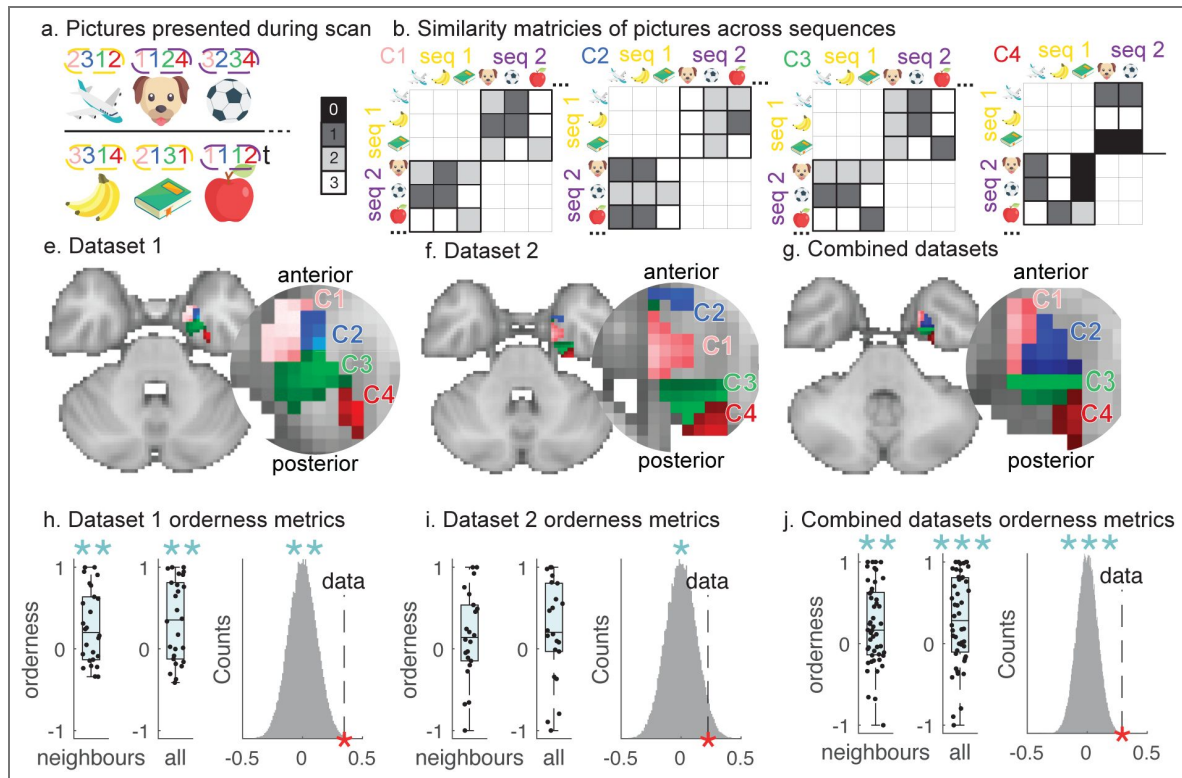


Fig 4. Abstract generalisable coordinate system.

a. We presented pictures associated to specific coordinates in random order in the MRI scanner. Notice that some were paired with a sequence 1 coordinate during training (yellow dashed line) and others with sequence 2 (purple dashed line). **b.** Across-sequence similarity matrices. To find coordinate representations, we specify hypothesised similarities between pictures with respect to each level of the hierarchy. This time we include only similarities between pictures from different auditory sequences for further analysis. **c-e.** Visualisation of group averaged coordinate representations in dataset 1 (**c**), dataset 2 (**d**), and both combined (**e**). **f-h.** Orderliness metric for neighbouring levels pairs only (left), including all pairs of levels (middle), and averaged across participants against a null distribution from shuffled levels' labels (right) in dataset 1 (**f**), dataset 2 (**g**), and both combined (**h**).

Previous studies have described gradients of temporal and spatial granularity outside of entorhinal cortex, highlighting a broader principle of hierarchical organization in the human brain^{20–24}. However, these studies are looking for place-cell-like representations that vary in their resolution. In contrast, our study was designed to look for a factorial hierarchy (like the grid code), that repeats with different periodicities at different levels of the hierarchy. In principle it would be possible to check for other brain regions displaying the same factored hierarchy, although given the nature of the gradient analysis, this would require strong hypotheses about the gradient direction and boundaries in each anatomical location. Indeed, we were unable to find such a hierarchy even within right EC. Unilateral effects are common in fMRI studies of grid-like coding^{25–27}. While the reason for our left lateralization remains unclear, the replication across two independent datasets supports the robustness of the effect.

Previous studies have found representations of time in the medial temporal lobe. Time cells that respond to intervals between events exist in rodents^{8,9,28–31} and humans³². These include codes that are hierarchical, with ramping responses over multiple timescales, but not abstract, because they are tethered to sensory experience⁸. Others are abstract but not hierarchical, with transient responses that tile a single timescale^{28–32}. In our studies, such a representation would account for the data only if the representations of time were restructured according to the structure of events. In the cells, where the structure is linear, this means representations that tile time from the start to the end of each trial. In the fMRI, it would require an even larger restructuring, with time being measured independently for every level in hierarchy, with each level perfectly matching the same level in the auditory sequence. It is already known that temporal codes in the rodent lateral entorhinal cortex are indeed reshaped by simple event structures^{8,9}. However, so far this has been described for laplace-like representations that peak at the start and end of an event. Our cells do not have this form, as many peak for intermediate points in the sequence (as do previous coordinate cells for spatial sequences^{33,34}).

But our experiences are not distributed uniformly in space and time. They form richly structured sequences. Statistical dependencies in experience, for example sequences of goals in space^{26,33,34} and graph structures embedded in time^{10,35–37}, do indeed influence both fMRI signals and cellular responses in the human entorhinal cortex. These responses have been suggested to support the formation of a cognitive map: an internal representation that allows animals to reason about future outcomes. Learning a cognitive map requires understanding the relational structure that underlies experience and representing it in a way that facilitates flexible behaviour. Our study shows that such representations for general sequences beyond space and time can be abstracted to an efficient coordinate that compresses the experiential structure.

Strikingly, this hierarchical compression aligns with the anatomical gradient of the entorhinal cortex in exactly the fashion predicted by the grid cell hierarchy, though our data cannot identify the cellular substrate of this gradient, which could involve grid cells in the medial entorhinal cortex, time and task cells in the lateral entorhinal cortex, a combination of these, or other entorhinal populations. This raises the question of whether the two hierarchies are encoded by the same cells or whether there are separate abstract coordinate systems across domains. It is not possible to know this from our study. Recent theoretical work offers reasons to favour both possibilities¹¹. Forming a bespoke coordinate system that captures the experiential statistics allows efficient inferences and predictions about unseen events^{7,12,38} and efficient memory search³⁹. By contrast, in similarly structured models, using grid cells themselves as the coordinate scaffold for arbitrary non-spatial sequences provides stable memories that gracefully degrade at capacity limits¹¹. Testing between these ideas will likely require studies that measure representations of space and structured non-spatial sequences in the same cells. However, in either case, our data establishes that the anatomical hierarchy in the entorhinal cortex is present in humans and broadens its classical understanding beyond 2-dimensional spatial sequences to arbitrary sequences of experiences.

Methods

Intracranial data: participants

Intracranial recordings were acquired from 6 patients with pharmacologically resistant epilepsy in the intensive monitoring unit of the Klinik Lengg in Zurich, Switzerland. The ages of the participants ranged from 29 to 56 years (mean = 38.6, SD = 11.6). As part of their evaluation for neurosurgical epilepsy treatment, all patients had sEEG depth electrodes intracranial electrodes implanted in the medial temporal lobe for diagnostic purposes. During the experiment, no clinical seizures occurred. Informed consent was given by all patients. All participants provided written informed consent for the study, which was approved by the institutional ethics review board (PB 2016-02055). All participants had normal or corrected-to-normal vision.

Intracranial data: the task

The patients had to memorise a simple four-picture sequence. The same pictures were shown in every trial, but their order was randomised. Each image was shown for 1200 milliseconds at the center of the screen with no interstimulus interval. There was a 3-second delay after the sequence was presented. Following that, one of the pictures was presented, and the patients had to answer where it was in the sequence (position 1, 2, 3, or 4). Before the next trial, auditory feedback indicated whether the given response was correct. Each session included 50 trials, and most patients completed multiple sessions of the task (mean = 3.3, SD = 1.2).

Intracranial data: recordings setup

The depth electrodes (1.3 mm diameter, 8 contacts of 1.6 mm length, and spacing between contact centres of 5 mm; Ad-Tech, Racine, WI, www.adtechmedical.com) were stereotactically implanted into amygdala, hippocampus, and entorhinal cortex (Fig. S1.1a). In this work, we only focused on the recordings from entorhinal cortex and hippocampus (Fig S1.1b). Each macroelectrode had nine microelectrodes that protruded approximately 4 mm from its tip. The data from the microelectrodes was acquired at sampling frequencies of 32 kHz via the ATLAS recording system (0.5- to 5000-Hz passband, Neuralynx, Bozeman, MT, USA; www.neuralynx.com). The recording setup is described in detail in^{40,41}.

Anatomical localisation of implanted electrodes was performed with Lead-DBS, a specialized software for electrode mapping⁴². This process involved co-registration of pre-implantation MRI and post-implantation CT, normalization of pre-implantation MRI to MNI space, subcortical refinement to account for brain shift and manual annotation of electrode contact positions. Additionally, each patient's pre-implantation MRI was segmented with FreeSurfer⁴³, by applying cortical parcellation and subcortical structure segmentation to precisely determine anatomical electrode locations. The automated segmentation was verified, with a clinician's assessment serving as the final determination of electrode localization. The position of the microwire channels was assigned to the location of the first macro contact.

Spike sorting was performed using the Wave Clus software⁴⁴, followed by manual refinement to ensure data quality. The first multi-unit cluster was excluded to ensure we have only single-unit data. Clusters in which spikes were predominantly observed within noisy segments of the recording were removed, as were those exhibiting non-physiological noise or high proportion of inter-spike intervals (ISI) shorter than 3 ms. Additionally, clusters with highly similar waveforms were merged to prevent over-clustering (Fig S1.2).

Task structure

To explore whether neurons in the human hippocampal formation encode the structure of a task, we first visualised the averaged population trajectory throughout a full trial of the task in neural activity space. We calculated the firing rate of each unit in 10 ms bins, except during the initiation and question stage. These stages of the task had a variable duration across trials because they only end when the patient responds. To align these stages across trials, we divided them into a fixed

number of 50 (initiation) and 150 (question) bins in each trial, so that the bin size may vary from trial to trial. We then averaged the firing rate across trials for each unit so that we ended up with a $N \times T$ matrix of activity, where N is the number of cells and T the number of time bins. We did a principal component analysis and projected the data on the first three principal components to get a $3 \times T$ matrix that described the trajectory in neural space after dimensionality reduction. We smoothed this trajectory with a gaussian kernel (using matlab's `smoothdata`) of 25 bins (250 ms) before plotting (Fig 1c [↗](#)).

Since different task stages occupied different regions in neural activity space, we expected that task stages could be separated by their patterns of activity. To test this, we decoded task stage from the population firing rates using a Bayesian decoder. We calculated tuning curves for task stages by averaging the firing rate within each task stage across all odd trials, and then calculated the decoded posterior probability of each stage given the averaged firing rate within that stage for all even trials. We then repeated this procedure the other way around, decoding the task stage from the average of odd trials using tuning curves from averaging even trials. We averaged these two repeats, then plotted the resulting probability of decoding each task stage given the firing rates of each task stage (Fig 1d [↗](#) left). To calculate the decoding error, we subtracted the identity matrix from the decoding probabilities (Fig 1d [↗](#) right). We also plotted raster plots with overlaid average firing rate (after gaussian smoothing with a 50-bin kernel) across trials for two example cells that exhibit strong task structure coding (Fig 1e, f [↗](#)).

Sequence position

The stimulus presentation stages of the task provided the cleanest separation of structure and sensory input, because the stimuli were randomised across positions across trials. In each trial, we presented the same four stimuli, but in a random order, so that responses to stimulus identity and sequence position could be identified separately. For each stimulus and each sequence position, we collected all occurrences where that stimulus appeared in that position. We averaged firing rates across these occurrences for each 10 ms time bin for each neuron. We then smoothed the firing rates with a gaussian kernel (using matlab's `smoothdata`) of 30 bins (300 ms) to obtain a neural response through time for the given stimulus in the given position for each neuron (Fig 1h-k [↗](#)). Additionally, for a more compact representation of the same data, we first averaged the firing rate through time in each stimulus-in-position occurrence, and then averaged across occurrences, to obtain a 4×4 (position by stimulus) matrix of average firing rates (Fig 1l [↗](#)).

We then asked whether either of these variables, i.e. stimulus identity and sequence position, significantly modulated the firing rate of individual neurons. To do so, we averaged each unit's firing rate across time for each stimulus presentation in each trial, to get a $4 \times K$ -dimensional vector with K the number of trials and ran a two-factor Anova (type III sum of squares) where one factor is the stimulus identity and the other the sequence position. We then counted the number of units that come out significant for either or both (Fig 1m [↗](#)).

Finally, we calculated the coefficient of partial determination for stimulus identity and sequence position across the neural population. We started from the $4 \times K$ -dimensional vector of stimulus responses at each timepoint for each unit. We then fitted generalised linear models that modelled stimulus identity and sequence position responses with three regressors each. All three stimulus regressors contained a -1 for when stimulus 1 was presented, and each a +1 for stimulus 2, 3, and 4 respectively; similarly, all position regressors contained a -1 for position 1, and each a +1 for position 2, 3, and 4 respectively. We calculated the sum of squared residuals both for the full model including stimulus and position regressors, and a partial model without the regressors for one of those. The coefficient of partial determination was given by the difference between the sum of squared residuals of the partial and the full model, divided by those of the partial model. After calculating the coefficient of partial determination for each timepoint for each unit, we averaged across units to end up with one curve for stimulus identity and one for sequence position, then smoothed with a gaussian kernel of 30 bins (Fig 1n [↗](#)). In order to assess significance, we repeated

the exact same procedure but shuffled the labels for stimulus position and identity and built a distribution of the maximum coefficient of partial determination through time for each permutation. The significance level was given by the 95% percentile of that distribution.

fMRI experiments: participants

We carried out two fMRI experiments with the same task but two independent cohorts of participants. 30 healthy participants participated in each study (dataset 1: 30 participants aged 20–30, mean = 23.7, SD = 2.90); dataset 2: 30 participants aged 19–35, mean = 23.8, SD = 3.6). We excluded the participants that did not reach the training criteria on the second day of the experiment (see below for an explanation of the day 2 training criteria; 2 participants were excluded from dataset 1 and 4 participants from dataset 2). We then analysed data only from the participants who drew the structure of auditory sequences correctly on day 4 (see below for an explanation of the day 4 testing criteria; 2 participants were excluded from dataset 1 and 4 participants from dataset 2). Hence, all the analysis presented is based on data from 26 participants in dataset 1 and 22 participants in dataset 2. Monetary compensation was £15 per hour of fMRI scan, with 20p for each catch trial in the scanner and £10 per hour for the behavioural tasks. The study was approved by the local ethics committee (ethics number R43594/RE001). Each individual signed an informed consent form prior to the start of the experiment.

fMRI experiments: auditory sequence design and behavioural training

To mimic the hierarchical structure of real-life experience of time (meals make up days, days make up weeks, and weeks make up months), we designed two auditory sequences (Fig 2f,g and Fig S2.1) with an analogous four-level hierarchical structure. Crucially, each sequence's sounds were different (Fig 2g and Fig S2.1). This allowed us to ask whether the neural representation of the structure generalised across different sensory stimuli. Both sequences were 65.5 seconds long in total. Each sound lasted for half a second. We familiarised our participants extensively with the auditory sequences by behavioural training over multiple days (Fig 2e).

Day 1. The participants received audio recordings of both sequences on the first day. To ensure that participants paid attention to the recordings (as training was conducted at home on this day), we asked them to count the number of times a sequence fragment appeared in each recording. The behavioural training took 30 minutes on average, during which each participant heard each sequence 28 times.

Day 2. On the second day, we trained the participants until they learned the auditory sequences very well. The behavioural training consisted of two different tasks, and the training criterion was 100% accuracy on both tasks. The participants' first task required memorising the order of individual sounds in each sequence. Participants were presented with a sound and asked to determine which of three subsequent sounds was the next one in the sequence. Participants were given three seconds to respond, and each sound was presented for half a second. The second task emphasised on learning the sequence structure. Participants were asked to determine whether a test sequence was identical to the actual sequence ('correct') or whether it had been altered ('wrong'). There were six test sequences in total, three of which were identical to the original sequence and three altered in one of the levels. Alternations in a level were created by changing the number of repetitions in that level compared to the original sequence. Each training block consisted of participants hearing the true sequence three times and then performing both tasks.

Day 3. On the third day, we added pictures to the audio sequences. Each auditory sequence had eight pictures, each of which was presented in a unique location in the sequence at the same time as the sound it was paired with (0.5 s). We excluded sounds that indicated boundaries (sound E,F,G/T,U,V in Fig 2g) for pairing with pictures. Picture locations were randomly distributed across the participants. We used 4 objects in 4 different orientations in the first experiment to make 16 pictures, and 16 different objects in the second experiment. This manipulation of the orientations of pictures was designed to ask a question in the first dataset which we do not

investigate in this manuscript. While passively watching and listening to the audio sequence, participants were asked to remember the location of each visual stimulus in the sequence. Each sequence was played for the participants 25 times.

Day 4. Similarly to day 3, participants were presented with each of the two sequences ten times prior to the scan task to refresh their memories of them. We then acquired fMRI measurements while projecting the pictures at the center of a screen in a pseudorandomized order, without sounds. Thanks to this order randomisation in the scanner we avoid autocorrelation confounds while we measure the coordinate representation associated with the picture on the screen. As a cover task, we instructed participants to press a button whenever a picture was stretched horizontally or vertically (the distractor picture). Every picture was presented twenty times in its original form and once as a distractor image with different sizes. Each participant completed 4 blocks of the task, with 336 trials in each block. The pictures were shown for 0.8 seconds on each trial. Each visual stimulus was presented with a jittered inter-trial interval (ITI) selected from a truncated gamma distribution with a mean of 2.5 s.

After the fMRI scan task, we tested how well the participants remembered the location of each visual stimulus in the audio sequence. First, participants were asked to draw the structure of sequences 1 and 2 with pen and paper. They were then asked to locate the position of each of the 16 different pictures in their drawing of the sequences (for an example drawing see [Supplementary Fig 2.2](#)).

fMRI experiments: behavioural analysis

Custom MATLAB scripts (Mathworks, Version R2018/b) were used for all statistical analysis. To determine the accuracy of each picture placement in each level by each participant, we examined the participants' post-scan drawings. We calculated their accuracy by comparing the positions of each of the 16 pictures in the participant's drawing to the coordinate of those pictures in each hierarchical level of the true sequence. The percentage accuracy in each level was compared to chance in a one-tailed t-test to determine whether participants remembered the locations of the pictures in the auditory sequences. The chance levels for levels 1-3 were 33% and 25% for level 4.

Then, we analysed the participants' errors. We calculated the distance from the correct location for each incorrectly placed picture. These error distances are an indication of whether the participants exploited the hierarchical sequence structure to remember picture positions, through an increased likelihood of hierarchical errors over just linear proximity. In the days-weeks-months analogy, this is equivalent to misremembering that you had dinner with friends on Wednesday, a week later, instead of the immediately following Saturday—despite the latter being closer in time. We counted the number of errors for each error distance and regressed those on the linear error distance and the hierarchical error distance in a multiple regression. We defined the hierarchical error distance as the Euclidian distance between the true four-dimensional hierarchical coordinate and the erroneous four-dimensional coordinate. This regression thus reported to what extent the linear distance and the hierarchical distance explain the number of errors across participants. To determine if the hierarchical distance predicted errors better than the linear distance, we did a one-sided t-test on the contrast of parameter estimates between the two.

fMRI experiments: fMRI data acquisition

We presented visual stimuli on a screen in the scanner. MRI data were collected from a 64-channel head coil on a 3-Tesla Siemens Magnetom Prisma MRI scanner. A multi-band sequence was used with a multi-band factor of 3 (MB3). Slicing was done with an in-plane resolution of 2×2mm with no breaks and a slice thickness of 2mm each. The slices were acquired in interleaved order, with an echo time (TE) of 20 ms and a repetition time (TR) of 1.235 s.

The data were collected in four separate blocks per participant. The number of acquired measurements varied per block, with an average of 867 volumes per block.

A field map of dual-echo time with whole brain coverage was collected with a voxel size of 2 × 2 × 2 mm³ (1st TE: 4.92 ms, 2nd TE: 7.38 ms). T1-weighted structural images were collected for each participant for co-registration with echo-planar images (EPI). Each T1-weighted image is

constructed of 192 axial slices of 1mm each, in-plane resolution = 1×1 mm², with the following parameters: field of view (FoV) = 192 mm, TE = 3.97ms and TR = 1.90s.

fMRI experiments: fMRI data preprocessing

Pre-processing was performed using tools from the fMRI Expert Analysis Tool (FEAT), part of FMRIB's Software Library (FSL⁴⁵). Data for each of the 8 blocks was preprocessed separately. Each block was aligned to the first, pre-saturated image using the motion-correction tool MCFLIRT⁴⁶. Brain extraction was performed using the automated brain extraction tool BET⁴⁷. All data were high pass temporally filtered with a cut-off of 100 s. Registration of EPI images to high-resolution structural images and to standard (MNI) space was performed using FMRIB's Linear and Non-Linear Registration Tool (FLIRT⁴⁶ and FNIRT⁴⁸), respectively. The registration transformations were then used to map each block's EPI data to the native structural space. No spatial smoothing was performed during pre-processing.

fMRI experiments: RSA

Using representational similarity analysis (RSA¹⁷), we identified how the brain represented the coordinates of the four levels of auditory sequences. We created four independent hypothesis representational dissimilarity matrices (RDMs) by calculating the distances between all the pictures in all four levels, as each image had a coordinate in all four levels of the hierarchy.

The RSA pipeline was built on the RSA toolbox¹⁸. The preprocessed data were first transformed from functional to MNI standard space. By defining a mask from the Harvard-Oxford Subcortical Structural Atlas in fsleyes, only cerebral cortical and hippocampal voxels were included. Within this mask, we defined 100-voxel searchlights for each voxel. Then, we calculated the average stimulus responses in each searchlight by fitting a general linear model (GLM) to the BOLD data of every voxel within the searchlight. This GLM modelled every one of the four task blocks separately. Each block contained onset regressors for each of the 16 images, as well as separate regressors for catch trials and button presses. We also added motion outlier frames and the average signal in a CSF region of interest as nuisance regressors. Except for nuisance regressors, SPM12 (Wellcome Trust Centre for Neuroimaging, <https://www.fil.ion.ucl.ac.uk/spm>) convolves regressors with a hemodynamic response function and fits the GLM to produce a searchlight activation pattern: the average response to a stimulus for each voxel in the searchlight. Given the activation pattern for each stimulus in each block, we calculated across-block RDMs $X^{AB}_{ij} = 1 - \text{mean}(\text{corr}(x_i^A, x_j^B), \text{corr}(x_j^A, x_i^B))$ where x_y^Z is the activation pattern x for stimulus y in block Z ; the mean ensures this dissimilarity matrix is symmetric. We then averaged the across-block RDM X^{AB} across all combinations of blocks where $A \neq B$. By excluding any similarities between pictures within the same block, we avoid biased statistics due to regressor correlations⁴⁹. As a result, we end up with one 16×16 data RDM, containing dissimilarities between each pair of the 16 pictures. As the last step in the within-participant RSA pipeline, we used multiple regression where we regressed the upper triangle of the data RDM on the upper triangle of each of the four hypothesis RDMs (one for each coordinate) in every searchlight. Alternatively, the four hypothesis RDMs could be regressed separately; this produces similar results (Fig S3.5, Supplementary Table 5). The resulting parameter maps were then smoothed with a 5 mm full width at half maximum kernel.

The result of the above-described analyses is a set of brain maps containing regression coefficients in each voxel for each participant. Then, using fMRI Expert Analysis Tool (FEAT)⁴⁵, we calculate statistical maps at the group level to assess the reliability of these results across the entire participant population⁴⁵. The GLM at the group level contained a mean effect regressor.

fMRI experiments: orderliness metric

Next, we asked if the representations for each level were arranged in a gradient along the anterior-posterior dimension of each participant's entorhinal cortex. To do this, we first identified the location of the centre of mass (CoM) of each level's effect within the left entorhinal cortex (parallel analyses in right entorhinal cortex yielded a null result: Fig S3.2, Table 2), defined by

the Juelich atlas⁵⁰ anatomical mask thresholded at 50% (results for a stricter 75% mask threshold are reported in Fig S3.3 [↗](#), Supplementary Table 3 [↗](#)). We extracted each coordinate CoM's location along the A-P axis, yielding a single number for coordinate i referred to as Y_i . Then, we determined if the Y values were consistent with an ordered distribution, more so than chance. To quantify to what extent the A-P locations were distributed along the hypothesised order, we defined an 'orderness metric'.

The orderness metric was defined as follows: $\frac{\sum_{i,j>i}^{\text{pairs}} (Y_i - Y_j)}{\sum_{i,j>i}^{\text{pairs}} |Y_i - Y_j|}$. Importantly, for each pair of coordinates included in this measure, we always subtracted Y for the coordinate with the *higher* index from the Y for the coordinate with the *lower* index. To be consistent with our hypothesised order, this difference should be positive, because 1) we have defined Y to increase in the anterior direction of the A-P axis, and 2) the lower-index (lower frequency) coordinate is expected to be more anterior than the higher-index (higher frequency) coordinate. Because we divided by the absolute value of the difference, this metric was normalised between -1 and 1 where 1 indicated perfect ordering, so $Y_1 Y_2 Y_3 Y_4$ in A-P order, and -1 opposite ordering, so $Y_4 Y_3 Y_2 Y_1$ in A-P order. We separated contributions to the orderness metric from neighbour coordinates versus all coordinates by calculating it twice: once with only neighbouring pairs, i.e. $([1,2], [2,3], [3,4])$, and once with all pairs, i.e. $([1,2], [1,3], [1,4], [2,3], [2,4], [3,4])$. Finally, we also calculated a discrete version of the orderness metric, that incorporates the sign of the difference rather than the continuous difference value: (results in Fig S3.4 [↗](#), Supplementary Table 4 [↗](#)).

We calculated the orderness metric for each participant, both including only neighbouring pairs and including all pairs of coordinates. To evaluate whether the distribution of coordinate representations was significantly ordered as hypothesised, we tested whether the orderness metric was significantly positive across participants. First, we evaluated the significance of a one-tailed t-test against zero of orderness metrics across the population. Then, we constructed a null distribution of population orderness metrics by randomly permuting the coordinate labels for each participant, recalculating the orderness metric for the permuted labels, and averaging the orderness metric across the population. We produced population orderness metrics for 100000 of these permutations. We calculated the permutation test p-value as the fraction of permutations with a larger orderness metric than the population-averaged orderness metric with unpermuted coordinate labels.

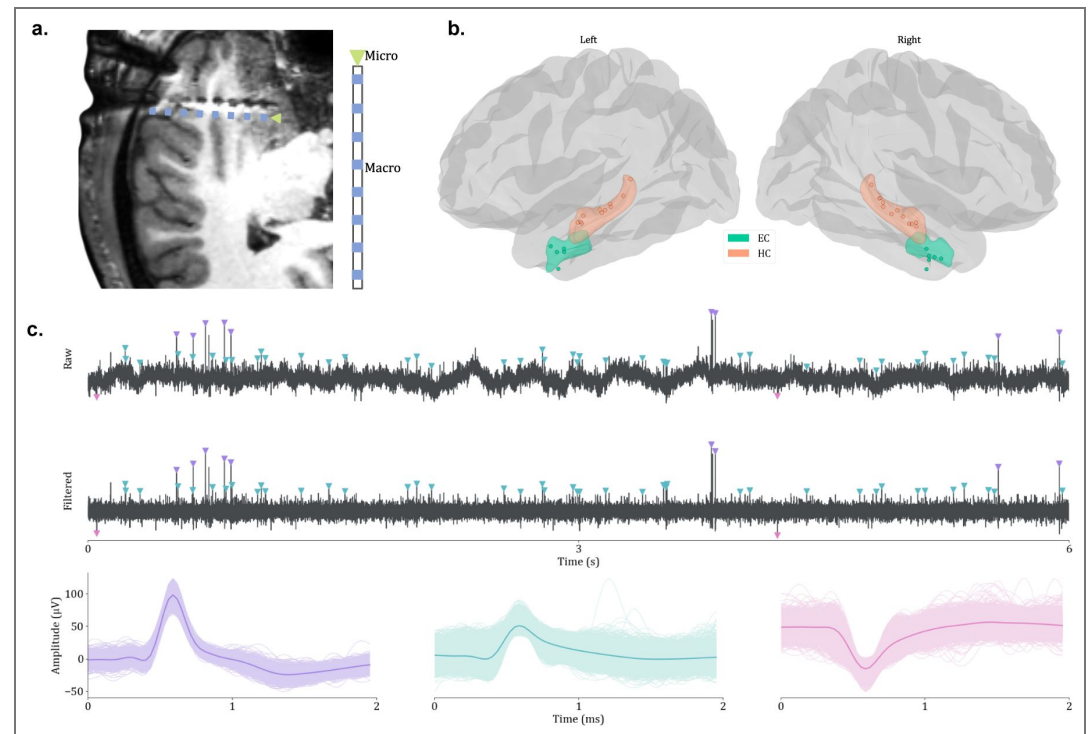
fMRI experiments: ROI analysis

We next tested if the centres of mass (CoM) of each level's coordinate representation were in the same anatomical locations in entorhinal cortex across the two datasets. We first used the group maps to identify the CoM of each level in the first dataset. We then calculated the effect of each level for each participant in the second dataset at the CoMs defined from the first dataset. We then repeated the procedure but now defined the CoM from the second dataset and tested the effects in the first dataset. We average these two 4×4 matrices together to get a 4×4 matrix E_{ij} containing the effect of level i in dataset 1 at the CoM of level j in dataset 2, averaged with the effect of level i in dataset 2 at the CoM of level j in dataset 1, for each participant. We then compared whether congruent effects were stronger than incongruent ones (e.g., the effect of the third level at the peak of the second level defined in the different dataset vs. the peak of the second level). To do so, we performed a one-tailed paired t-test across participants between the averages of the diagonal elements (congruent effects) and the averages of the off-diagonal elements (incongruent effects).

fMRI experiments: generalisation analysis

Finally, we tested if the coordinate representations generalised across the two sequences, which had the same structure but different sounds. We again calculated hypothesis RDMs but now the distances for each level were defined only based on the distances between pictures in different sequences (Fig 4b [↗](#)). We then performed the same orderness metric analysis as described above. We also performed the same analysis for the hypothesis RDMs based only on the distances between the pictures within the same sequences.

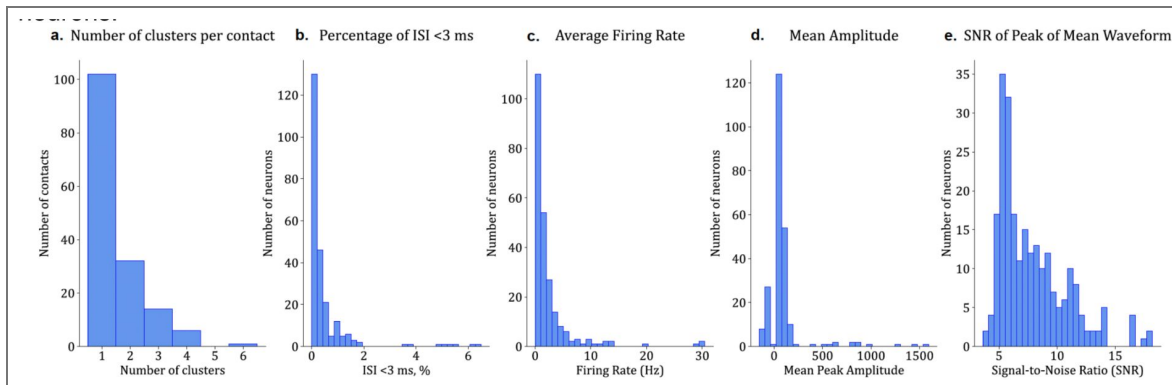
Supplementary Material



Supplementary figure 1.1. a. Combinatorial electrode. Right: schematic showing macro- (blue) and microelectrode (green) contacts along the depth electrode. Left: example MRI from a patient illustrates electrode placement, with visible macro- and microcontacts, with an overlaid schematic of the electrode positioned slightly below its actual location in the MRI. b. Electrode localization in MNI space, showing electrode positions in the Freesurfer average brain model, with the entorhinal cortex (EC, green) and hippocampus (HC, orange) highlighted. Electrodes are color-coded according to their respective anatomical regions. c. Spike sorting example, showing raw recordings (top) and recordings filtered in 300-3000 Hz range (middle) from a single microelectrode channel. Detected spikes are marked and color-coded corresponding to the three identified neurons (bottom). Individual spike waveforms (lighter color) and averaged waveforms (darker color) are displayed for three identified neurons.

Supplementary figure 1.2. Spike Sorting Quality Assessment.

Summary of key metrics used to evaluate the reliability of spike sorting: **a.** the number of neuronal units identified per electrode contact, **b.** the percentage of inter-spike intervals (ISI) shorter than 3 ms, **c.** the average firing rate in Hz **d.** the peak amplitude (μV) of detected spikes, calculated as the mean across waveform peaks for identified neurons, and **e.** the signal-to-noise ratio (SNR) of the peak of the mean waveform, defined as the ratio of the peak amplitude of the neuron's mean spike waveform to the background noise level (calculated as the peak of the averaged waveform divided by the mean of the standard deviations of individual spike waveforms).



Supplementary Table 1. P-values table for Fig 3.

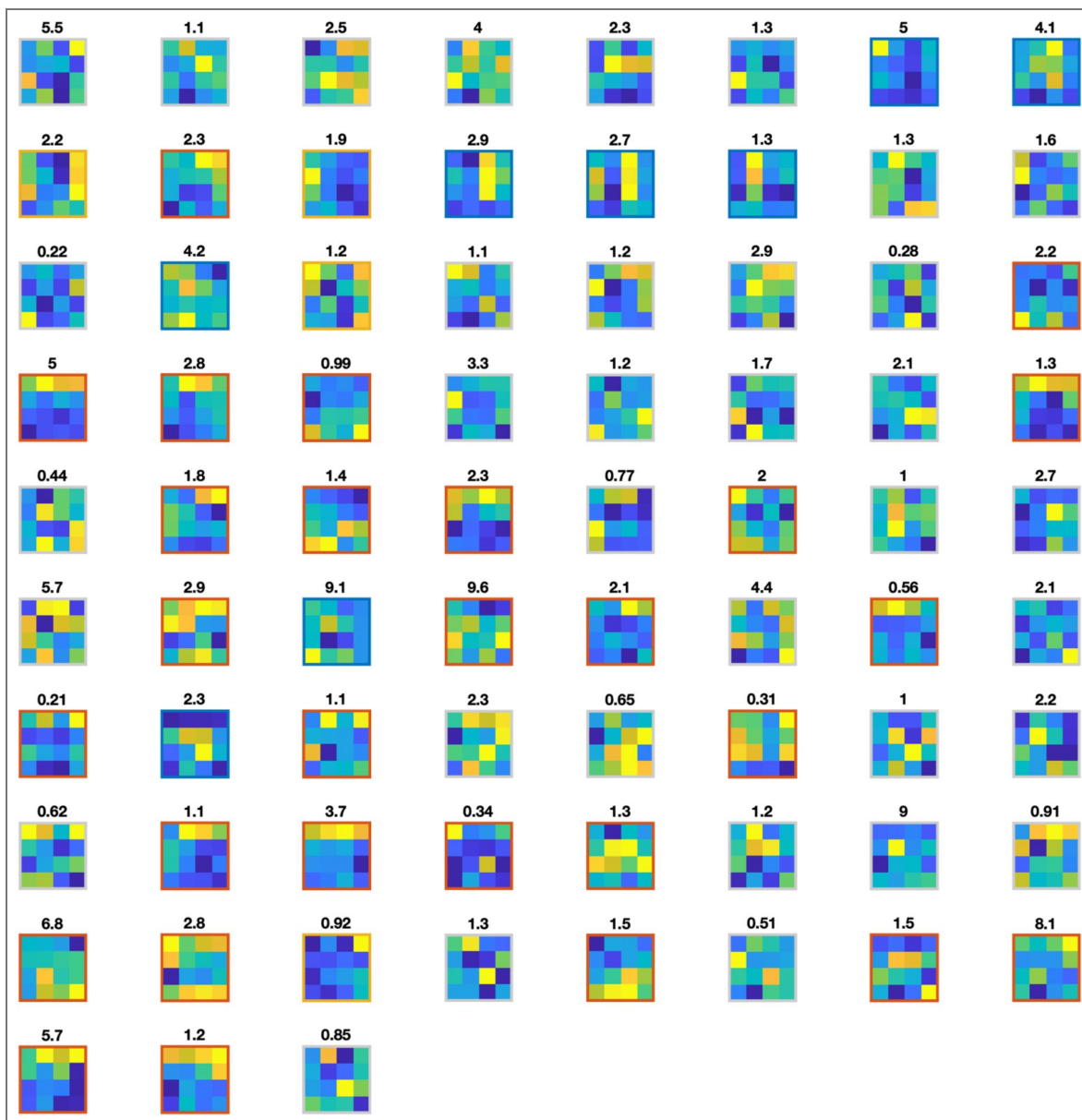
P-values for orderness metric for neighbouring pairs of levels (left column), including all pairs of levels (middle column), and averaged across participants against a null distribution from shuffled level labels (right column) in dataset 1 (first row), dataset 2 (second row), and both combined (third row). The rows correspond to figure 3h-j: first row to 3h, second row to 3i, third row to 3j. The columns also correspond to figure 3h-j: first column to 3h-j left, the second column to 3h-j middle and the third column to 3h-j right. For Dataset 1, the significance threshold was corrected for multiple comparisons across both hemispheres.

patient	sessions	hpc	hpc_pos	hpc_pic	ec	ec_pos	ec_pic
3	1	4	3	0	0	0	0
4	4	63	33	16	23	7	9
5	3	4	2	0	7	4	0
6	4	4	1	1	44	24	3
7	3	17	9	1	1	0	0



Supplementary figure 1.3. Summary firing rate plots for all hippocampal cells.

For each cell we average firing rates across time and across trial stages with a particular picture (columns) in a particular position (rows), then plot the 4×4 firing rate map with the maximum firing rate in yellow and the minimum firing rate in blue. The number above each plot indicates the maximum averaged firing rate, i.e. the firing rate for the yellow square (if that number is 0, it means the cell didn't fire during the stimulus presentation stages of the task, but it was active in different task stages). The box outline denotes whether the cell is significantly tuned to picture and position (blue), position only (orange), picture only (yellow), or neither (grey).

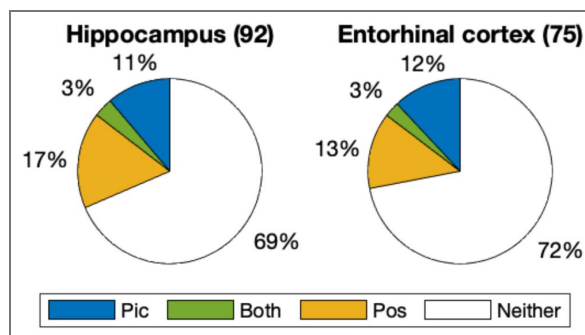


Supplementary figure 1.4. Summary firing rate plots for all entorhinal cells.

For each cell we average firing rates across time and across trial stages with a particular picture (columns) in a particular position (rows), then plot the 4×4 firing rate map with the maximum firing rate in yellow and the minimum firing rate in blue. The number above each plot indicates the maximum averaged firing rate, i.e. the firing rate for the yellow square (if that number is 0, it means the cell didn't fire during the stimulus presentation stages of the task, but it was active in different task stages). The box outline denotes whether the cell is significantly tuned to picture and position (blue), position only (orange), picture only (yellow), or neither (grey).

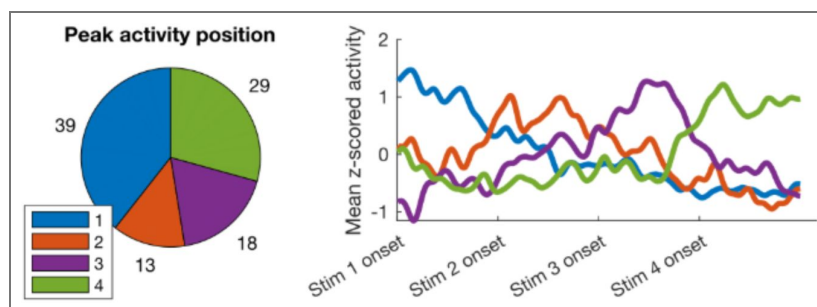
Supplementary figure 1.5. Single unit tuning to only position 2 and 3.

Cells that significantly code for stimulus position, picture, or both (analogous to Fig 1m) when excluding all data from position 1 and 4.



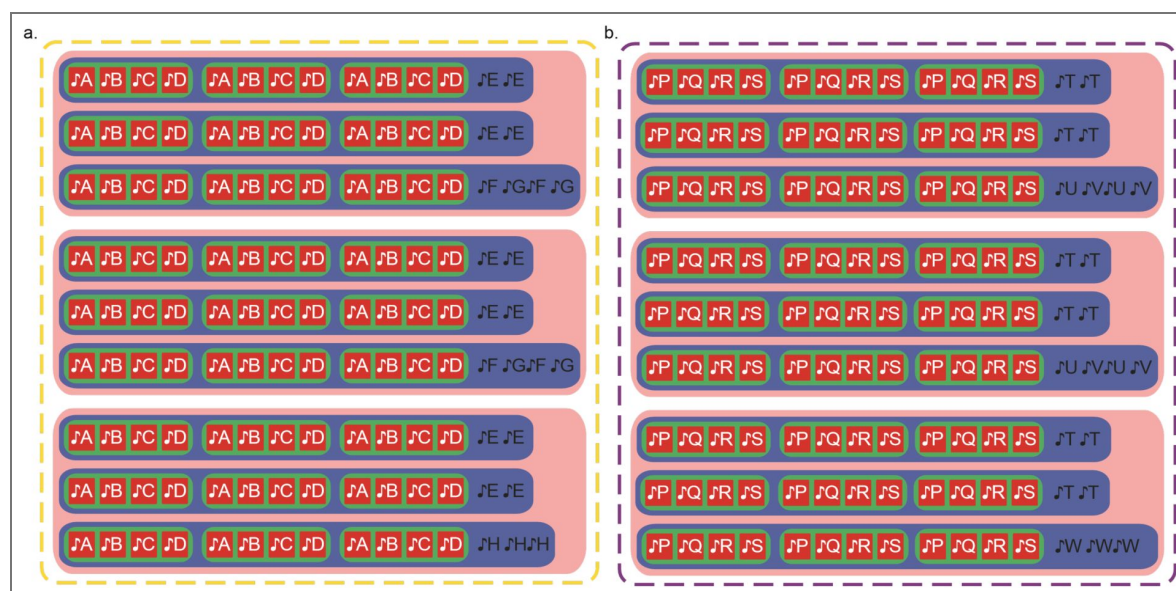
Supplementary figure 1.6. Single unit position preference.

Number of cells within position coding cells that have a maximum firing rate for each position (left), with the average z-scored firing rates of each group plotted throughout the stimulus presentations stages of the trial (right).



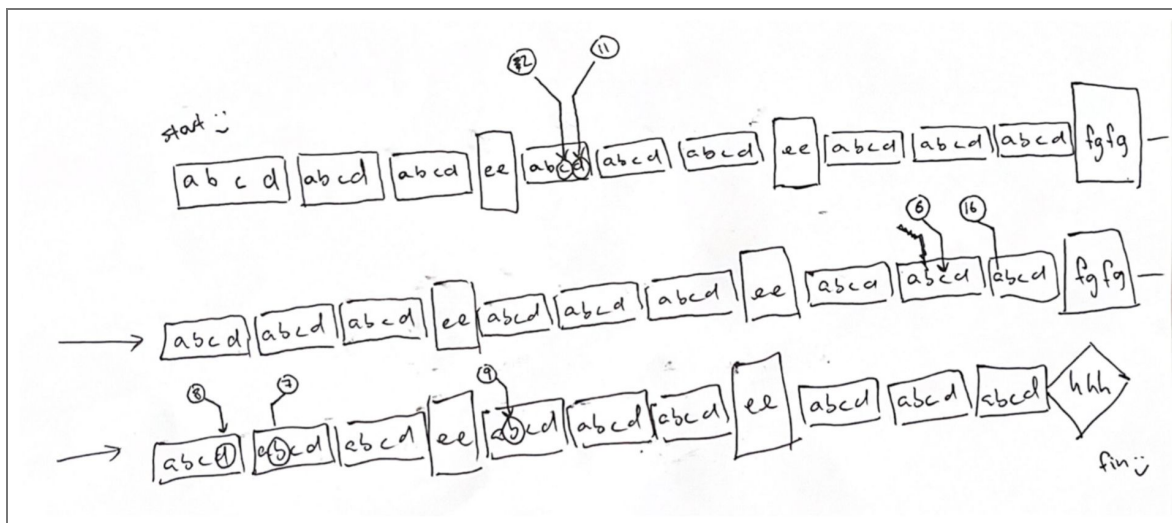
Supplementary figure 2.1. Auditory sequences.

Participants learnt two sequences with identical structures but different sounds.



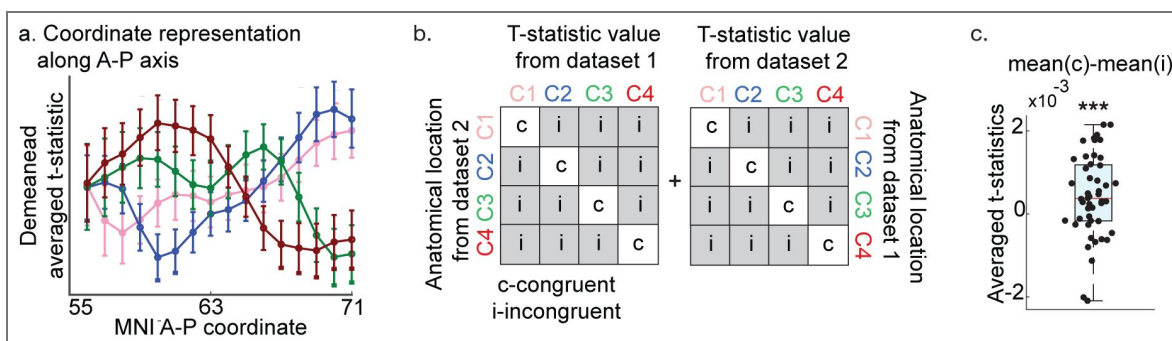
Supplementary figure S2.2. An example participant drawing of the auditory sequence and pictures in it.

The letters are indicating the sounds in the sequence, and the numbers are referring to the pictures. We did not explicitly instruct participants about the hierarchical nature of the sequence; the division along blocks and rows is their own interpretation.



Supplementary figure S3.1.

a. T-statistic for each level's coordinate representation (C1 is pink, C2 is blue, C3 is green, C4 is red) averaged across participants within A-P slices as a function of slice A-P position in MNI space. Note that the peaks are ordered according to the hypothesised order (both datasets together). b. To test for the anatomical consistency of EC coordinate representations across datasets, we compare a level's t-statistic at that level's centre of mass from the other dataset (c: congruent) against its t-statistics at other levels centre of mass from the other dataset (i: incongruent). See Methods for more details. c. Congruent effects were stronger than non-congruent effects, $p=0.0039$.



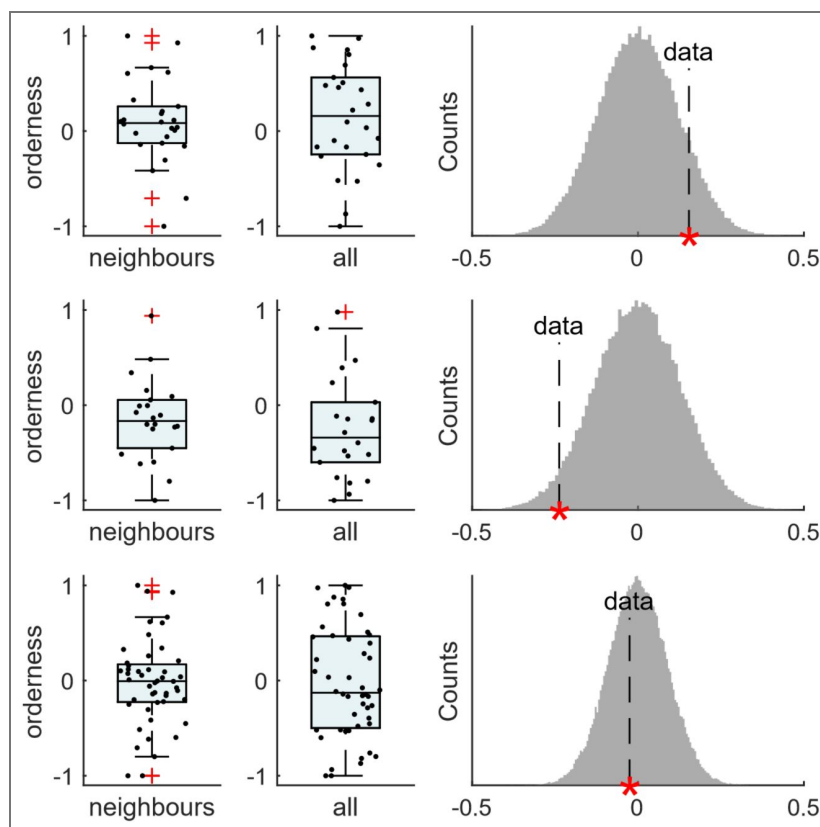
Supplementary Table 1. Cells and sessions per patient.

Number of sessions completed for each patient, with the result number of hippocampal (hpc) and entorhinal (ec) units and the corresponding neurons significantly tuned to position (hpc_tbls1).

Data	Neighbours	All	Permutations
Dataset 1	0.00065	0.00002	0.00028
Dataset 2	0.01911	0.04002	0.02291
Both	0.00009	0.00003	0.00004

Supplementary figure S3.2. The orderness metrics computed on centres of mass defined in the right entorhinal mask.

Orderness metric for neighbouring pairs of levels (left column), including all pairs of levels (middle column), and averaged across participants against a null distribution from shuffled level labels (right column) in dataset 1 (first row), dataset 2 (second row), and both combined (third row). See corresponding p-values in the table below.



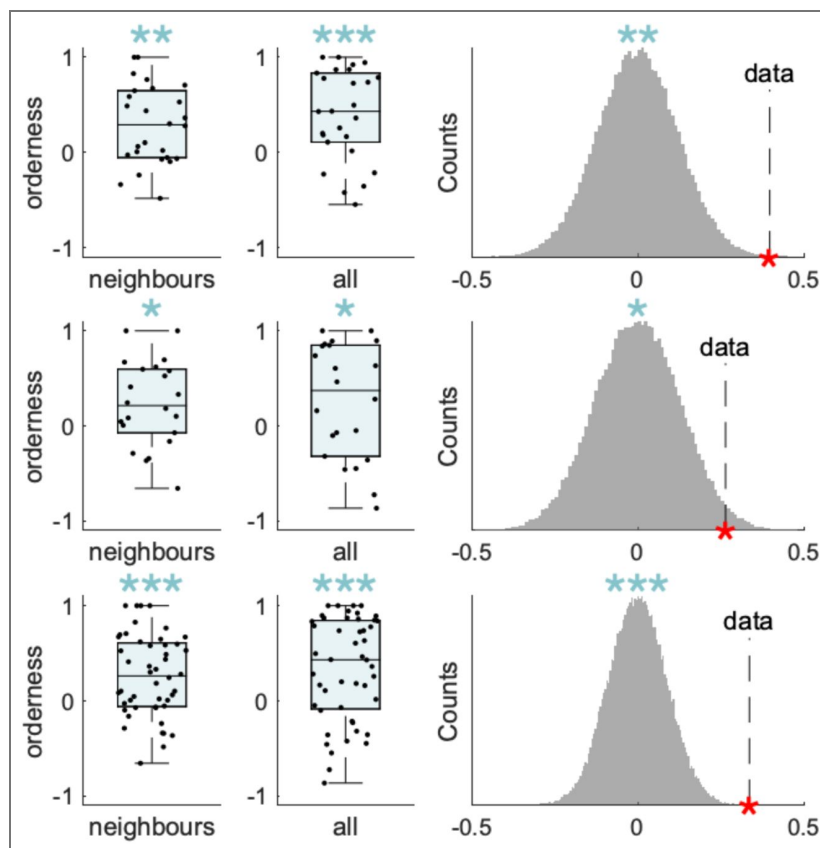
Supplementary Table 2. P-values table, corresponding to the Fig S3.2, for the orderness metrics computed on centres of mass defined in the right entorhinal mask.

For Dataset 1, the significance threshold was corrected for multiple comparisons across both hemispheres.

Data	Neighbours	All	Permutations
Dataset 1	0.14596	0.0883	0.104
Dataset 2	0.94396	0.97322	0.96731
Both	0.6126	0.61871	0.61262

Supplementary figure S3.3. The orderness metrics computed on centres of mass defined in the stricter entorhinal mask (atlas probability 75% instead of the 50% probability mask used in all other analyses, see Methods).

Orderness metric for neighbouring pairs of levels (left column), including all pairs of levels (middle column), and averaged across participants against a null distribution from shuffled level labels (right column) in dataset 1 (first row), dataset 2 (second row), and both combined (third row). See corresponding p-values in the table below.



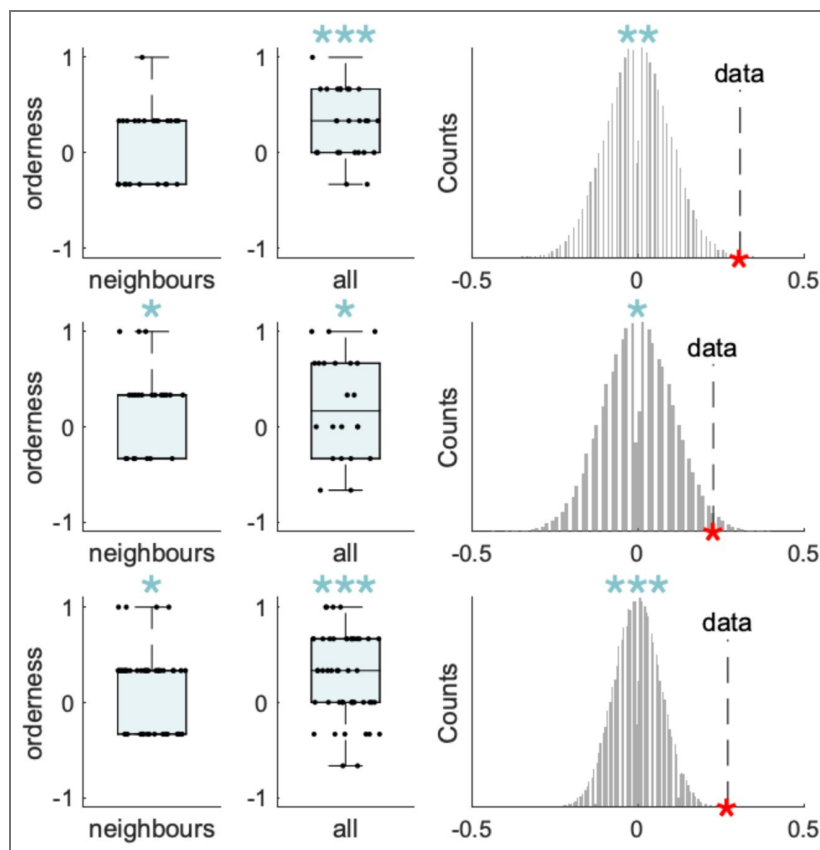
Supplementary Table 3. P-values table, corresponding to the Fig S3.3, for the orderness metrics computed on centres of mass in the stricter entorhinal mask.

For Dataset 1, the significance threshold was corrected for multiple comparisons across both hemispheres.

Data	Neighbours	All	Permutations
Dataset 1	0.00093	0.00014	0.00042
Dataset 2	0.01131	0.02752	0.02018
Both	0.00005	0.00004	0.00007

Supplementary figure S3.4. Discrete orderness metric

(see Methods) for neighbouring pairs of levels (left column), including all pairs of levels (middle column), and averaged across participants against a null distribution from shuffled level labels (right column) in dataset 1 (first row), dataset 2 (second row), and both combined (third row). See corresponding p-values in the table below.



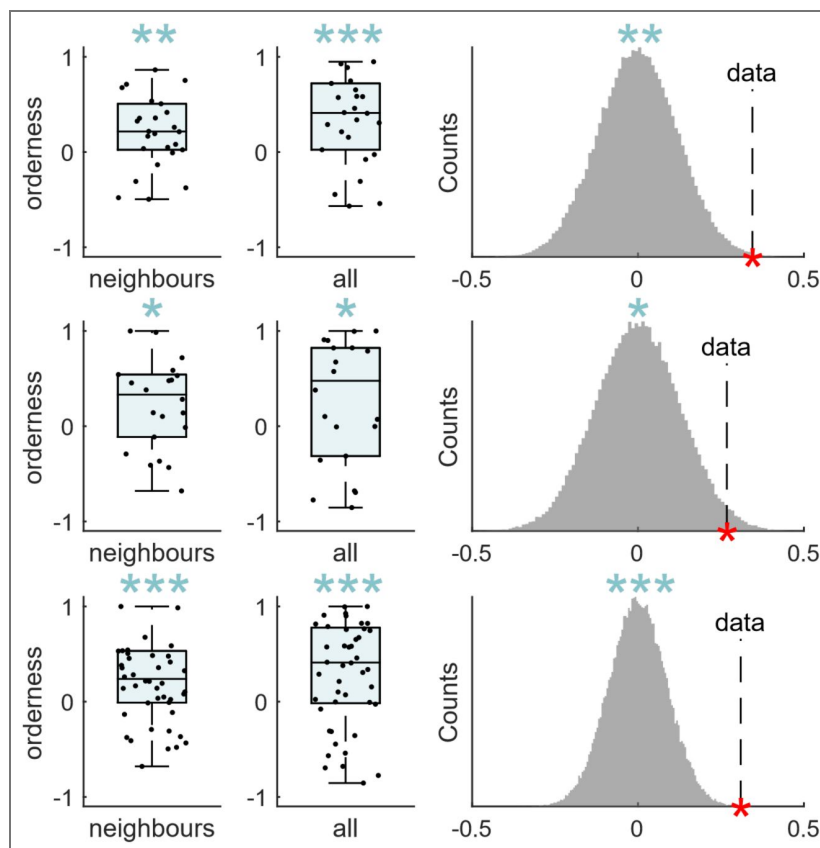
Supplementary Table 4. P-values table, corresponding to the Fig S3.4, for the discrete orderness metric.

For Dataset 1, the significance threshold was corrected for multiple comparisons across both hemispheres.

Data	Neighbours	All	Permutations
Dataset 1	0.15639	0.00008	0.00041
Dataset 2	0.0178	0.03047	0.01219
Both	0.01185	0.00005	0.00005

Supplementary figure S3.5. The orderness metrics computed on centres of mass defined on statistical maps that result from separate regressions (fitting 4 similarity regressors separately (one for each level), Methods).

Orderness metric for neighbouring pairs of levels (left column), including all pairs of levels (middle column), and averaged across participants against a null distribution from shuffled level labels (right column) in dataset 1 (first row), dataset 2 (second row), and both combined (third row). See corresponding p-values in the table below.



Supplementary Table 5. P-values table, corresponding to the Fig S3.5, for the orderness metrics computed on statistical maps that result from a multiple regression.

For Dataset 1, the significance threshold was corrected for multiple comparisons across both hemispheres.

	Neighbours	All	Permutations
Dataset 1	0.00355	0.00037	0.00173
Dataset 2	0.01312	0.03182	0.01935
Both	0.00023	0.00014	0.00015

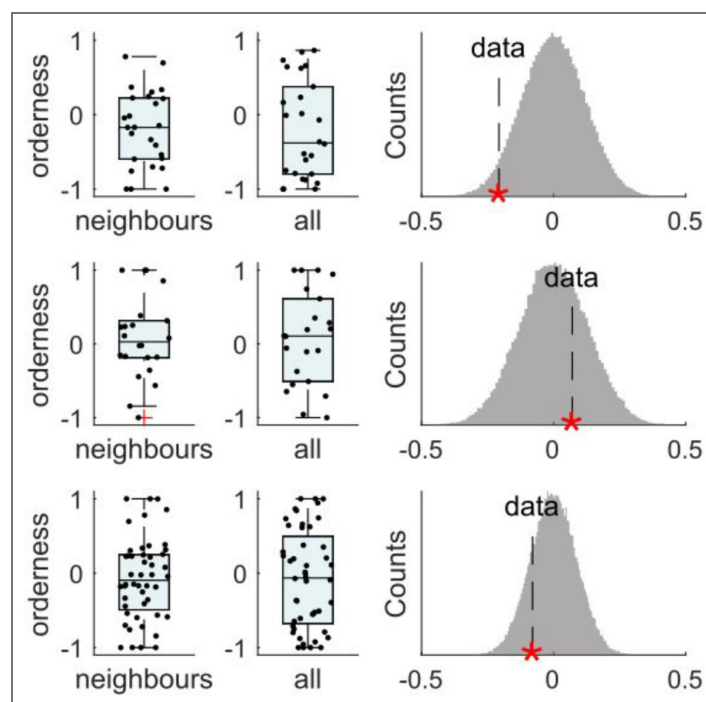
Supplementary Table 6. P-values table for Fig 4: across-sequence similarities matrices.

P-values for orderness metric for neighbouring pairs of levels (left column), including all pairs of levels (middle column), and averaged across participants against a null distribution from shuffled level labels (right column) in dataset 1 (first row), dataset 2 (second row), and both combined (third row). The rows correspond to Fig 4f-h: first row to 3f, second row to 3g, third row to 3h. The columns also correspond to Fig 3f-h: first column to 3f-h left, the second column to 3f-h middle and the third column to 3f-h right. For Dataset 1, the significance threshold was corrected for multiple comparisons across both hemispheres.

Data	Neighbours	All	Permutations
Dataset 1	0.00256	0.0008	0.00171
Dataset 2	0.09224	0.05094	0.04066
Both	0.00171	0.00034	0.00036

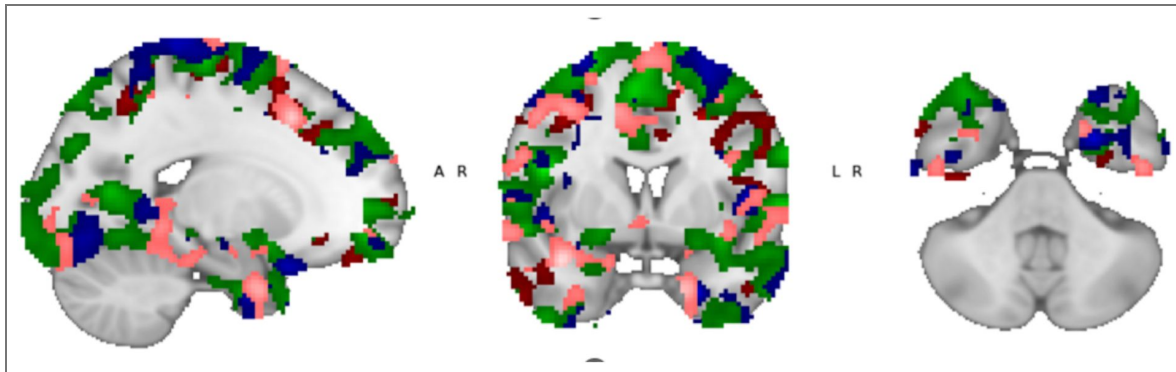
Supplementary figure. 3.6. The null result for orderness metrics computed on centres of mass defined in the hippocampal mask.

Orderness metric for neighbouring pairs of levels (left column), including all pairs of levels (middle column), and averaged across participants against a null distribution from shuffled level labels (right column) in dataset 1 (first row), dataset 2 (second row), and both combined (third row).



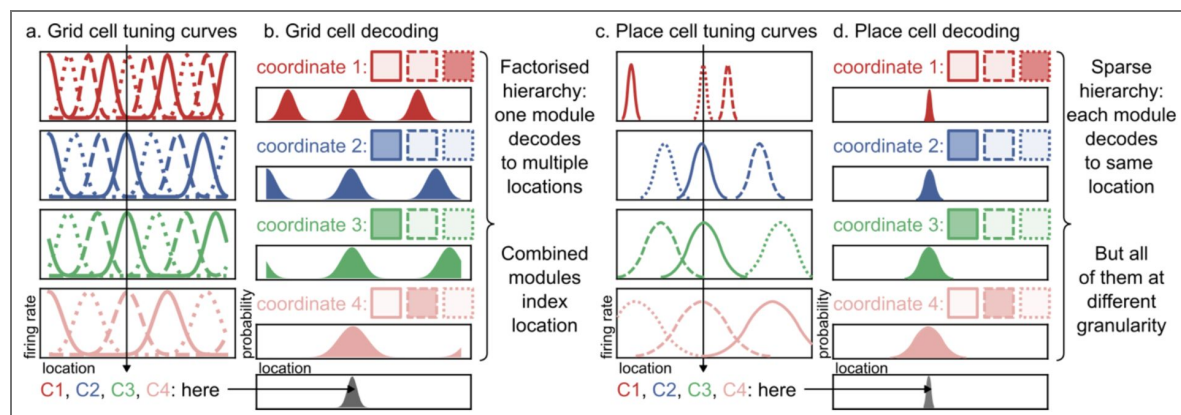
Supplementary figure. 3.7. Whole-brain statistical maps for the representational similarity analysis (RSA) of each hierarchical layer coordinate.

Maps are shown at a low statistical threshold for visualization purposes. Note that these group-level maps were not used for any statistical analysis. Instead, we use subject-level statistical maps to define a within-subject measure of representational order (the “orderness metric”), which is the reported statistic in all analyses. This within-subject order measure is more sensitive because it avoids signal loss due to averaging in a small brain region that is difficult to align across participants. Because widespread activation is typical in whole-brain fMRI analyses, our study used a strictly hypothesis-driven approach, focusing specifically on representational order within subjects within the entorhinal cortex.



Supplementary figure 5. Grid cell and place cell hierarchies.

a. Grid cell tuning curves, exhibiting periodic responses at different phases within a module (dotted/dashed/solid lines) at different scales across modules (colours). b. Decoded location from each module independently (colours) and jointly (grey, bottom row). These grid responses implement a factorised hierarchy, where each module encodes an independent hierarchical coordinate, and decoding a single location is only possible by combining the representations across modules – just like a calendar event is only fully determined by time, date, and month and not by any one of those. c. Place cell tuning curves that show individual peaks at different locations (dotted/dashed/solid lines) at different spatial scales (colours). d. Decoded location from each module independently (colours) and jointly (grey, bottom row). The location can be decoded separately from every module, at different levels of granularity – just like a street address that specifies street, town, and country.



Data availability

Data and analysis code supporting the findings of this study will be deposited in a public repository (Open Science Framework) upon publication. The single-unit and fMRI datasets contain human neural recordings; access will be provided in accordance with participant consent and ethics approval. Analysis code is available from the corresponding author and will be made publicly available on GitHub upon publication.

Acknowledgements

We thank the participants for their engagement with this difficult task. We thank Salman E Qasim, Joshua Jacobs, Timothy H Muller and Yunzhe Liu for their insightful comments and practical help. We also thank Emile Radyte and Tess Van Stekelenburg for help in the analysis process.

A.S. acknowledges a Wellcome Trust studentship and NIH funding. T.E.J.B. is supported by a Wellcome Principal Research Fellowship (219525/Z/19/Z), a Wellcome Collaborator award (214314/Z/18/Z), and by the Jean Francois and Marie-Laure de Clermont Tonerre Foundation. The Wellcome Centre for Integrative Neuroimaging and Wellcome Centre for Human Neuroimaging are each supported by core funding from the Wellcome Trust (203139/Z/16/Z, 203147/Z/16/Z). The Sainsbury-Wellcome centre is supported by core funding from the Wellcome Trust (219627/Z/19/Z) and the Gatsby Charitable Foundation (GAT3755) and the Gatsby Initiative for Brain Development and Psychiatry (GAT3955).

Additional information

Author contribution

Conceptualization and experimental design: A.S., H.B., T.E.J.B. Single unit data collection: A.S., E.M.S., J.S., D.L., L.I. fMRI data collection: A.S. and E.M.S. Data analysis: A.S., J.J.W.B., A.K., H.C.B., T.E.J.B. Manuscript writing: A.S., J.J.W.B., A.B.B., T.E.J.B. Manuscript editing: A.S., J.J.W.B., A.B.B., J.S., D.L., L.I., A.K., E.M.S., H.C.B., T.E.J.B.

Funding

Funder	Grant reference number	Author
Wellcome Trust (WT)		Anna Shpektor
Wellcome Trust (WT)	https://doi.org/10.35802/219525	Timothy E Behrens

Author ORCID iDs

Anna Shpektor:  <https://orcid.org/0000-0002-1554-5674>

Johannes Sarnthein:  <https://orcid.org/0000-0001-9141-381X>

Timothy EJ Behrens:  <https://orcid.org/0000-0003-0048-1177>

References

1. **Hafting T**, Fyhn M, Molden S, Moser M.-B, Moser EI (2005) Microstructure of a spatial map in the entorhinal cortex. *Nature* **436**:801-806 <https://doi.org/10.1038/nature03721> | [PubMed](#)
2. **Stensola H**, et al. (2012) The entorhinal grid map is discretized. *Nature* **492**:72-78 <https://doi.org/10.1038/nature11649> | [PubMed](#)
3. **Stemmler M**, Mathis A, Herz AVM (2015) Connecting multiple spatial scales to decode the population activity of grid cells. *Sci. Adv* **1** <https://doi.org/10.1126/science.1500816> | [PubMed](#)
4. **Bush D**, Barry C, Manson D, Burgess N (2015) Using Grid Cells for Navigation. *Neuron* **87**:507-520 <https://doi.org/10.1016/j.neuron.2015.07.006> | [PubMed](#)

5. Fyhn M, Hafting T, Treves A, Moser M.-B, Moser EI (2007) Hippocampal remapping and grid realignment in entorhinal cortex. *Nature* **446**:190-194 <https://doi.org/10.1038/nature05601> | [PubMed](#)
6. Behrens TEJ, et al. (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> | [PubMed](#)
7. Whittington JCR, et al. (2020) The Tolman-Eichenbaum Machine: Unifying Space and Relational Memory through Generalization in the Hippocampal Formation. *Cell* **183**:1249-1263.e23 <https://doi.org/10.1016/j.cell.2020.10.024> | [PubMed](#)
8. Tsao A, et al. (2018) Integrating time from experience in the lateral entorhinal cortex. *Nature* **561**:57-62 <https://doi.org/10.1038/s41586-018-0459-6> | [PubMed](#)
9. Kanter BR, Lykken CM, Polti I, Moser MB, Moser EI (2025) Event structure sculpts neural population dynamics in the lateral entorhinal cortex. *Science* **388** <https://doi.org/10.1126/science.adr0927> | [PubMed](#)
10. Bellmund JL, Deuker L, Doeller CF (2019) Mapping sequence structure in the human lateral entorhinal cortex. *eLife* **8** <https://doi.org/10.7554/elife.45333> | [PubMed](#)
11. Chandra S, Sharma S, Chaudhuri R, Fiete I (2025) Episodic and associative memory from spatial scaffolds in the hippocampus. *Nature* <https://doi.org/10.1038/s41586-024-08392-y> | [PubMed](#)
12. Dordek Y, Soudry D, Meir R, Derdikman D (2016) Extracting grid cell characteristics from place cell inputs using non-negative principal component analysis. *eLife* **5**:e10094 <https://doi.org/10.7554/eLife.10094> | [PubMed](#)
13. Yoon K, Lewallen S, Kinkhabwala AA, Tank DW, Fiete IR (2016) Grid Cell Responses in 1D Environments Assessed as Slices through a 2D Lattice Article Grid Cell Responses in 1D Environments Assessed as Slices through a 2D Lattice. *Neuron* **89** <https://doi.org/10.1016/j.neuron.2016.01.039> | [PubMed](#)
14. Stachenfeld KL, Botvinick MM, Gershman SJ (2017) The hippocampus as a predictive map. *Nat. Neurosci* **20**:1643-1653 <https://doi.org/10.1038/nn.4650> | [PubMed](#)
15. Mark S, et al. (2024) Flexible neural representations of abstract structural knowledge in the human Entorhinal Cortex. *eLife* **13** <https://doi.org/10.7554/eLife.101134> | [PubMed](#)
16. Strange BA, Witter MP, Lein ES, Moser EI (2014) Functional organization of the hippocampal longitudinal axis. *Nat. Rev. Neurosci* **15**:655-669 <https://doi.org/10.1038/nrn3785> | [PubMed](#)
17. Kriegeskorte N, Mur M, Bandettini P (2008) Representational similarity analysis - connecting the branches of systems neuroscience. *Front. Syst. Neurosci* **2** <https://doi.org/10.3389/neuro.06.004.2008> | [PubMed](#)
18. Nili H, Wingfield C, Walther A, Su L, Marslen-Wilson W (2014) A Toolbox for Representational Similarity Analysis. *PLoS Comput Biol* **10** <https://doi.org/10.1371/journal.pcbi.1003553> | [PubMed](#)
19. Kriegeskorte N, Goebel R, Bandettini P (2006) Information-based functional brain mapping. *Proc. Natl. Acad. Sci. U. S. A* **103**:3863-3868 <https://doi.org/10.1073/pnas.0600244103> | [PubMed](#)
20. Collin SHP, Milivojevic B, Doeller CF (2015) Memory hierarchies map onto the hippocampal long axis in humans. *Nat. Neurosci* **18**:1562-1564 <https://doi.org/10.1038/nn.4138> | [PubMed](#)
21. Momennejad I (2024) Memory, Space, and Planning: Multiscale Predictive Representations. *arXiv* <https://doi.org/10.48550/arXiv.2401.09491>
22. Brunec IK, Momennejad I (2022) Predictive Representations in Hippocampal and Prefrontal Hierarchies. *J. Neurosci. Off. J. Soc. Neurosci* **42**:299-312 <https://doi.org/10.1523/jneurosci.1327-21.2021> | [PubMed](#)
23. Nielson DM, Smith TA, Sreekumar V, Dennis S, Sederberg PB (2015) Human hippocampus represents space and time during retrieval of real-world memories. *Proc. Natl. Acad. Sci. U. S. A* **112**:11078-11083 <https://doi.org/10.1073/pnas.1507104112> | [PubMed](#)

24. Liu C, Todorova R, Tang W, Oliva A, Fernandez-Ruiz A (2023) Associative and predictive hippocampal codes support memory-guided behaviors. *Science* **382** <https://doi.org/10.1126/science.adi8237> | PubMed
25. Constantinescu AO, O'Reilly JX, Behrens TEJ (2016) Organizing conceptual knowledge in humans with a gridlike code. *Science* **352** <https://doi.org/10.1126/science.aaf0941> | PubMed
26. Bellmund JLS, Deuker L, Schröder TN, Doeller CF (2016) Grid-cell representations in mental simulation. *eLife* **5** <https://doi.org/10.7554/elife.17089> | PubMed
27. Doeller CF, Barry C, Burgess N (2010) Evidence for grid cells in a human memory network. *Nature* **463**:657-661 <https://doi.org/10.1038/nature08704> | PubMed
28. Salz DM, et al. (2016) Time Cells in Hippocampal Area CA3. *J. Neurosci. Off. J. Soc. Neurosci* **36**:7476-7484 <https://doi.org/10.1523/jneurosci.0087-16.2016> | PubMed
29. Kraus BJ, Robinson RJ, White JA, Eichenbaum H, Hasselmo ME (2013) Hippocampal "Time Cells": Time versus Path Integration. *Neuron* **78**:1090-1101 <https://doi.org/10.1016/j.neuron.2013.04.015> | PubMed
30. MacDonald CJ, Carrow S, Place R, Eichenbaum H (2013) Distinct hippocampal time cell sequences represent odor memories in immobilized rats. *J. Neurosci. Off. J. Soc. Neurosci* **33**:14607-16 <https://doi.org/10.1523/jneurosci.1537-13.2013> | PubMed
31. Kraus BJ, et al. (2015) During Running in Place, Grid Cells Integrate Elapsed Time and Distance Run. *Neuron* **88**:578-589 <https://doi.org/10.1016/j.neuron.2015.09.031> | PubMed
32. Umbach G, et al. (2020) Time cells in the human hippocampus and entorhinal cortex support episodic memory. *Proc. Natl. Acad. Sci. U. S. A* **117**:28463-28474 <https://doi.org/10.1073/pnas.2013250117> | PubMed
33. Tsitsiklis M, et al. (2020) Single-Neuron Representations of Spatial Targets in Humans. *Curr. Biol* **30**:245-253.e4 <https://doi.org/10.1016/j.cub.2019.11.048> | PubMed
34. Khazali MF, et al. (2024) A preserved neural code for temporal order between memory formation and recall in the human medial temporal lobe. *bioRxiv* <https://doi.org/10.1101/2024.10.12.618011> | PubMed
35. Park SA, Miller DS, Boorman ED (2021) Inferences on a multidimensional social hierarchy use a grid-like code. *Nat. Neurosci* **24**:1292-1301 <https://doi.org/10.1038/s41593-021-00916-3> | PubMed
36. Garvert MM, Dolan RJ, Behrens TE (2017) A map of abstract relational knowledge in the human hippocampal-entorhinal cortex. *eLife* **6** <https://doi.org/10.7554/elife.17086> | PubMed
37. Tacikowski P, Berger CC, Ehrsson HH (2017) Dissociating the neural basis of conceptual self-awareness from perceptual awareness and unaware self-processing. *Cereb. Cortex* **1-14** <https://doi.org/10.1093/cercor/bhx004> | PubMed
38. Sreenivasan S, Fiete I (2011) Grid cells generate an analog error-correcting code for singularly precise neural computation. *Nat. Neurosci* **14**:1330-1337 <https://doi.org/10.1038/nn.2901> | PubMed
39. McNamee DC, Stachenfeld KL, Botvinick MM, Gershman SJ (2021) Flexible modulation of sequence generation in the entorhinal-hippocampal system. *Nat. Neurosci* **24**:851-862 <https://doi.org/10.1038/s41593-021-00831-7> | PubMed
40. Boran E, et al. (2019) Persistent hippocampal neural firing and hippocampal-cortical coupling predict verbal working memory load. *Sci. Adv* **5** <https://doi.org/10.1126/sciadv.aav3687> | PubMed
41. Boran E, Hilfiker P, Stieglitz L, Sarnthein J, Klaver P (2022) Persistent neuronal firing in the medial temporal lobe supports performance and workload of visual working memory in humans. *NeuroImage* **254** <https://doi.org/10.1016/j.neuroimage.2022.119123> | PubMed
42. Horn A, Kühn AA (2015) Lead-DBS: a toolbox for deep brain stimulation electrode localizations and visualizations. *NeuroImage* **107**:127-135 <https://doi.org/10.1016/j.neuroimage.2014.12.002> | PubMed

43. Fischl B (2012) FreeSurfer. *NeuroImage* **62**:774 <https://doi.org/10.1016/j.neuroimage.2012.01.021> | PubMed
44. Chaure FJ, Rey HG, Quiroga RQ (2018) A novel and fully automatic spike-sorting implementation with variable number of features. *J. Neurophysiol* **120**:1859-1871 <https://doi.org/10.1152/jn.00339.2018> | PubMed
45. Jenkinson M, Beckmann CF, Behrens TEJ, Woolrich MW, Smith SM (2012) FSL. *NeuroImage* **62**:782-790 <https://doi.org/10.1016/j.neuroimage.2011.09.015> | PubMed
46. Jenkinson M, Bannister P, Brady M, Smith S (2002) Improved optimization for the robust and accurate linear registration and motion correction of brain images. *NeuroImage* **17**:825-841 [https://doi.org/10.1016/s1053-8119\(02\)91132-8](https://doi.org/10.1016/s1053-8119(02)91132-8) | PubMed
47. Smith SM (2002) Fast robust automated brain extraction. *Hum. Brain Mapp* **17**:143-155 <https://doi.org/10.1002/hbm.10062> | PubMed
48. Jenkinson M, Smith S (2001) A global optimisation method for robust affine registration of brain images. *Med. Image Anal* **5**:143-56 [https://doi.org/10.1016/s1361-8415\(01\)00036-6](https://doi.org/10.1016/s1361-8415(01)00036-6) | PubMed
49. Walther A, et al. (2016) Reliability of dissimilarity measures for multi-voxel pattern analysis. *NeuroImage* **137**:188-200 <https://doi.org/10.1016/j.neuroimage.2015.12.012> | PubMed
50. Bludau S, et al. (2014) Cytoarchitecture, probability maps and functions of the human frontal pole. *NeuroImage* **93**:260-75 <https://doi.org/10.1016/j.neuroimage.2013.05.052> | PubMed

Peer reviews

Reviewer #1 (Public review):

Summary:

Shpektor et al. propose a link between how humans learn abstract and hierarchical structures to support memory (for example, remembering the event of the first landing on the moon) and the medial temporal lobe (MTL) and grid cells in particular. Given that there is solid work on how grid cells in different modules jointly encode position in rodents, providing evidence for the existence of a similar code in humans in the non-spatial domain and in relation to memory formation, would constitute a valuable finding.

The authors first examine a small human intracranial dataset to demonstrate that sequence position is decodable in MTL population codes. They then examine behavioral data from two larger groups of participants who passively viewed content presented in a hierarchical sequence and show that errors in recall of positions within that sequence qualitatively match hierarchical predictions. The task design enabled distinct signatures of memory representations at different levels of hierarchy. While there were no multivariate patterns in MTL or any brain region that matched these patterns reliably, a follow-up analysis in MTL revealed a gradient along the anterior-posterior axis, such that lower levels of the hierarchy tended to have representational peaks in more anterior regions of the MTL, which was consistent across the two fMRI datasets.

Major strengths of the study include the novelty of the experimental paradigm and data.

In particular, single cell recording in MTL from a small number of human participants during sequence learning and testing a larger group of human participants on a sequence amenable to hierarchical structure learning, and collecting fMRI data during retrieval.

Furthermore, the paper tackles an important question and does so from both directions, using inspirations from both biology and computational science to navigate it.

The primary weaknesses of the paper are a lack of compelling support for the overarching claim about hierarchical representation and a lack of clarity and consistency about exactly

what those hierarchical representations should and do look like. My concerns regarding these weaknesses are described below, and I believe that most, if not all, of them could be addressed through additional analysis and paper revisions.

In the first part of the paper, the authors provide single-cell recordings in MTL, and they report the existence of cells that are sensitive to position (more so than to picture). However, they don't elaborate on this result with a model for an abstract sequence code. This is an issue because one possible explanation for the sequential position decoding is that neurons just fire at the presentation of the first image and decay at different rates, or ramp up toward action or feedback. One might be able to decode the position in sequence from these cells' activity, but can hardly call this an abstract code of position in a sequence. However, the authors don't provide further investigation into what the single-cell result might suggest and move on to a completely different fMRI experiment in the second part of the paper. Being able to decode sequence position does not, in my view, necessarily imply an abstract positional code - and I felt that further analysis of the single unit data would be required to identify what representations gave rise to that decoding ability.

The most compelling evidence that participants were encoding temporal order hierarchically came from behavioral data in the second part of the paper. However, these results were not presented clearly enough to evaluate their reliability and specificity. Figure 2i shows histograms of errors across participants with arrows pointing to bars that apparently correspond to errors of different levels of hierarchy. There are three colored bars, corresponding to errors of one unit at the first, second, or third levels of hierarchy. The first level is not diagnostic of hierarchy, but the other two colored bars appear higher than the colors nearby them. However, my understanding is that these bars correspond to situations with the same tone - which seems like an obvious reason that two positions might be confused, which in my view would weaken the argument for hierarchical encoding. Furthermore, there is no display of variability in the plot or indication of individual differences, so it is hard to tell whether the histogram is dominated by a few participants who made a lot of errors or is reflective of a general tendency across participants.

The fMRI analyses, while creative, raise questions regarding interpretability. The authors report no representations of hierarchical position at any level, either in MTL or across the whole brain, which would typically be taken as a lack of evidence for the representations existing. Follow-up analyses revealed that what shadows of representations do exist seem to line up along the anterior-posterior gradient. But what does that mean if we can't be sure that the representations are really there? Typically, we tally up evidence supporting an overarching claim by testing multiple predictions that are all consistent with the same story - but in this case, it seems that not all such test results are consistent.

In many cases, it was difficult to judge the strength of evidence due to somewhat minimal reporting on the exact hypotheses tested and test statistics.

On a high level, I found the overarching story linking the two datasets together to be somewhat tenuous. While I understand that science rarely rolls out as a coherent story, presenting the authors' valuable experiments in this fashion makes it harder for the reader to digest the information and reach a conclusion. The relevance of the first section of the paper to the second is not immediately apparent. Each section provides somewhat incomplete evidence for a set of claims on its own - but my view was that combining the two studies led to more questions than answers - since the paradigms and measurements are so different.

In conclusion, the authors propose an interesting account of how memories are formed in the human brain, by building an abstract and hierarchical code. The paper identifies a few separate findings that are suggestive of hierarchical abstract memory encoding in the MTL - yet I believe that more work would need to be done to irrefutably support that claim.

<https://doi.org/10.7554/eLife.112067.1.sa2>

Reviewer #2 (Public review):

Overall, I think these are exciting results that make a very nice contribution to the literature. I thought the picture-tagging of sequence locations in the fMRI study was clever, and the across-sequence RSA results were especially compelling. But there are several aspects of the presentation of the results that reduced my confidence and enthusiasm.

(1) This is an unusual paper in that there is one human intracranial study and two fMRI studies. The paradigm for the intracranial study is very different than the fMRI paradigm. The key differences are that the fMRI paradigm is hierarchical, while the intracranial is flat, with no sequence learning component, and the fMRI is auditory, while the intracranial is auditory. The justification for the switch from intracranial to fMRI was that intracranial does not allow anterior-posterior axis analysis, but there are so many differences between the studies that this feels like an awkward transition and justification. Also, anterior-posterior analysis in the MTL may not be feasible in EC with intracranial data, but it can be feasible in the hippocampus, and indeed this could be very worthwhile and relevant to pursue (see point 2).

While the two independent fMRI datasets is a strength, the replications would have been much more compelling had the analysis for the second dataset been preregistered.

(2) The intracranial results are pitched as a novel "abstract coordinate representation" but there is a substantial prior literature on MTL "ordinal position codes", which I believe is the same thing in this paradigm. Most of this literature is in the hippocampus, which is, of course, very relevant given the hippocampal findings here, but there is also evidence for this kind of information in EC, e.g., <https://elifesciences.org/articles/45333> .

(3) Given the intracranial results in the hippocampus as well as the prior relevant literature on position coding, it was not clear why the hippocampus was not an ROI in the fMRI studies.

(4) It wasn't until reading the Methods section carefully that I understood that the results do not hold for the right EC, only the left. This deserves more acknowledgment.

(5) The use of one-sided t-tests with an alpha of .05 reduced my confidence in the robustness of the results.

<https://doi.org/10.7554/eLife.112067.1.sa1>

Reviewer #3 (Public review):

Summary:

Shpektor et al. investigate how hierarchical sequence structure is represented in the entorhinal cortex (EC) and medial temporal lobe (MTL) using a combination of single-unit recordings and fMRI. In the single-unit recordings, they find abstract representations of ordinal position within short sequences in both the EC and the hippocampus. Next, they use two fMRI datasets to examine representations of hierarchical sequence structure in EC. They find that these representations (1) are organized along a posterior-to-anterior hierarchy, with finer sequence structure represented in posterior EC and coarser structure in anterior EC, and (2) generalize across sensory features, suggesting an abstract representation of sequence position. The authors take these findings as evidence of a non-spatial hierarchical coordinate system in the human EC, analogous to grid cells in rodents.

Strengths:

The methodological approach presented in this study is commendable, combining single-unit recordings in the MTL with two fMRI datasets. The finding of hierarchical and abstract sequence representations in the EC is compelling and is replicated across these datasets and modalities. The manuscript addresses important questions about how the MTL abstracts across experiences that share hierarchical structure, a topic of considerable current interest. As such, the work is likely to be of broad interest to researchers studying these processes in both rodents and humans.

Weaknesses:

In my view, the main weaknesses concern the interpretation of the results, as well as several areas where additional analyses and methodological clarification would strengthen the manuscript. My point-by-point comments are as follows:

- (1) I found the evidence for hierarchical and abstract sequence-position representations interesting. However, I am less convinced by the stronger claim that these findings demonstrate a coordinate system analogous to grid-cell coding. The current results appear to provide stronger support for abstract sequence-position coding than for grid-like coding per se. In particular, it is not clear to me that hierarchical sequence representations necessarily imply a grid-like representational format or a coordinate system. Many neural systems exhibit gradients of representational scale along the anterior-posterior axis, both within and across brain regions, without being considered grid-like. I would encourage the authors to clarify why it should be interpreted specifically in terms of a coordinate system rather than more general hierarchical sequence representations. The manuscript would benefit either from a more explicit justification of this link to grid-cell coding or from a more cautious framing of the conclusions.
- (2) Relatedly, the emphasis on grid-cell-like coding naturally centers the story on entorhinal cortex (EC). Yet, the single-neuron results indicate that the hippocampus contained a comparable number of position-selective cells. In addition, a large body of literature has implicated the hippocampus in hierarchical representations of memories, sequences, and relational structure. For completeness, I encourage the authors to repeat the key fMRI analyses within the hippocampus, rather than focusing exclusively on EC.
- (3) I have some concerns regarding the amount of information available to distinguish representations at different levels of the sequence hierarchy. As I understand the design, each 113-tone sequence was associated with only eight images, meaning there were approximately 14 tones between successive image events. It would be helpful to provide additional detail regarding how image coordinates were assigned and selected, how many observations contributed to each hierarchical level, and how much statistical power was available to distinguish representations at different scales.
- (4) I was also uncertain about the potential influence of visual similarity in Dataset 1. My understanding is that the images were not entirely unique but instead consisted of rotated versions of the same images. If so, this visual similarity could potentially complicate the interpretation of representational structure. It would therefore be useful to clarify whether repeated images occurred within the same or different locations in the hierarchy and to provide analyses demonstrating that the reported effects cannot be explained by visual similarity. This seems particularly important given that the corresponding effects in Dataset 2 were weaker.
- (5) The rationale for using a custom orderliness metric could be explained more clearly. It would be helpful to understand why a custom metric was preferred over rank-order measures such as Kendall's tau or Spearman's rho. I would be interested in seeing whether the orderliness results replicate using one of these more conventional metrics.

(6) I had difficulty reconciling the finding that sequence representation effects are stronger across rather than within sequences. Intuitively, I would have expected representations within a sequence to reflect both shared hierarchical position and sensory experience, thus yielding stronger within-sequence effects than across sequences. The opposite pattern seems somewhat counterintuitive. I would appreciate additional discussion of this pattern and what it implies about the nature of the underlying representation. It would also be informative to know whether similar effects are observed elsewhere in the brain, and why EC might preferentially express a purely abstract representation more strongly than representations that additionally share sensory features.

(7) The authors' theory is that hierarchical representations of sequences in EC are used as a scaffold for memory, yet the current paper does not link their behavioural results to their neural ones. I think making such a link would greatly strengthen the results presented here. For example, is displacement error or sequence memory related to ordered representations of the sequence structure?

(8) I thought the manuscript would benefit from a broader discussion of prior work on (1) sequence representations and (2) hierarchical representations in the hippocampus and related regions. As it stands, the manuscript does a good job of situating its findings within the literature on grid cells in the EC but gives comparatively little attention to the literature on sequence representations in the hippocampus. Placing the current findings within this broader body of work would help clarify which aspects of the results are specific to a grid-like interpretation and which may instead reflect more general principles of hierarchical representation in the MTL or across the brain.

<https://doi.org/10.7554/eLife.112067.1.sa0>