

A hierarchical coordinate system for sequence memory in human entorhinal cortex

Anna Shpektor^{*1}, Jacob J. W. Bakermans^{*1,2}, Alon B. Baram¹, Johannes Sarnthein^{3,4}, Debora Ledergerber^{4,5}, Lukas Imbach^{4,5}, Emma Müller-Seydlitz¹, Helen C. Barron^{1,6}, Timothy E. J. Behrens^{1,7,8}

^{*}These authors contributed equally to this work

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

Grid cells in rodent entorhinal cortex (EC) support a coordinate system for space, enabling robust memory and powerful flexibility in spatial behaviour. This coordinate system is abstract - with the same grid cells encoding position across different sensory environments; and hierarchical - with grid modules of increasing spatial scale occupying increasingly ventral locations in the EC. Recent theories suggest that a similar abstract coordinate system could offer the same benefits to general memories that are not sequences drawn from a 2D surface. Here we show that an abstract hierarchical coordinate system supports arbitrary sequences in the human medial temporal lobe (MTL). In single-unit recordings from MTL, we find abstract, coordinate-like coding of a simple sequential memory task. In an fMRI experiment with more complex hierarchical sequences, we discover an abstract hierarchical representation in EC: the coordinate representations at distinct levels in the hierarchy are arranged on an anatomical gradient along the EC's anterior-posterior axis, homologous to the ventro-dorsal axis in rodents. These results therefore mirror the anatomical gradient of grid cells in the rodent EC but now for arbitrary non-spatial sequences. Together they suggest that memories are scaffolded on a hierarchical coordinate system using common neuronal coding principles, aligned to preserved anatomy, across domains and species.

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}. Recent research has argued grid cells also encode non-spatial and abstract 2D spaces^{8–12}, raising the possibility that the same neural system might function as a coordinate system for multiple 2D domains.

Here, we sought to test whether a similar neural coordinate system exists for the broader class of non-spatial sequences that do not lie on a 2D plane. Such a coordinate system could serve as a scaffold¹³ 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. 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 then use fMRI to probe a factorised, multiple coordinates representation of general hierarchical sequences, serving as a scaffold for memory. We hypothesised that in EC, this representation will have similar anatomical features to the EC's coordinate system for space – grid cells.

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. 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).

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 very high accuracy (Fig 1d, average posterior probability of correct label > 0.998). This representation was also present at the single neuron level: A one-way ANOVA revealed that most cells (Hippocampus: 77%, EC: 88%, Fig 1g) in the MTL were tuned to task stage (see example cells in Fig 1e,f and Figs S1.1, S1.2). Together, these results indicate that neurons in the MTL overwhelmingly represent different stages of the 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 sensory experience (e.g. common visual experience in all trials during the initiation or delay periods, and common auditory experience during feedback).

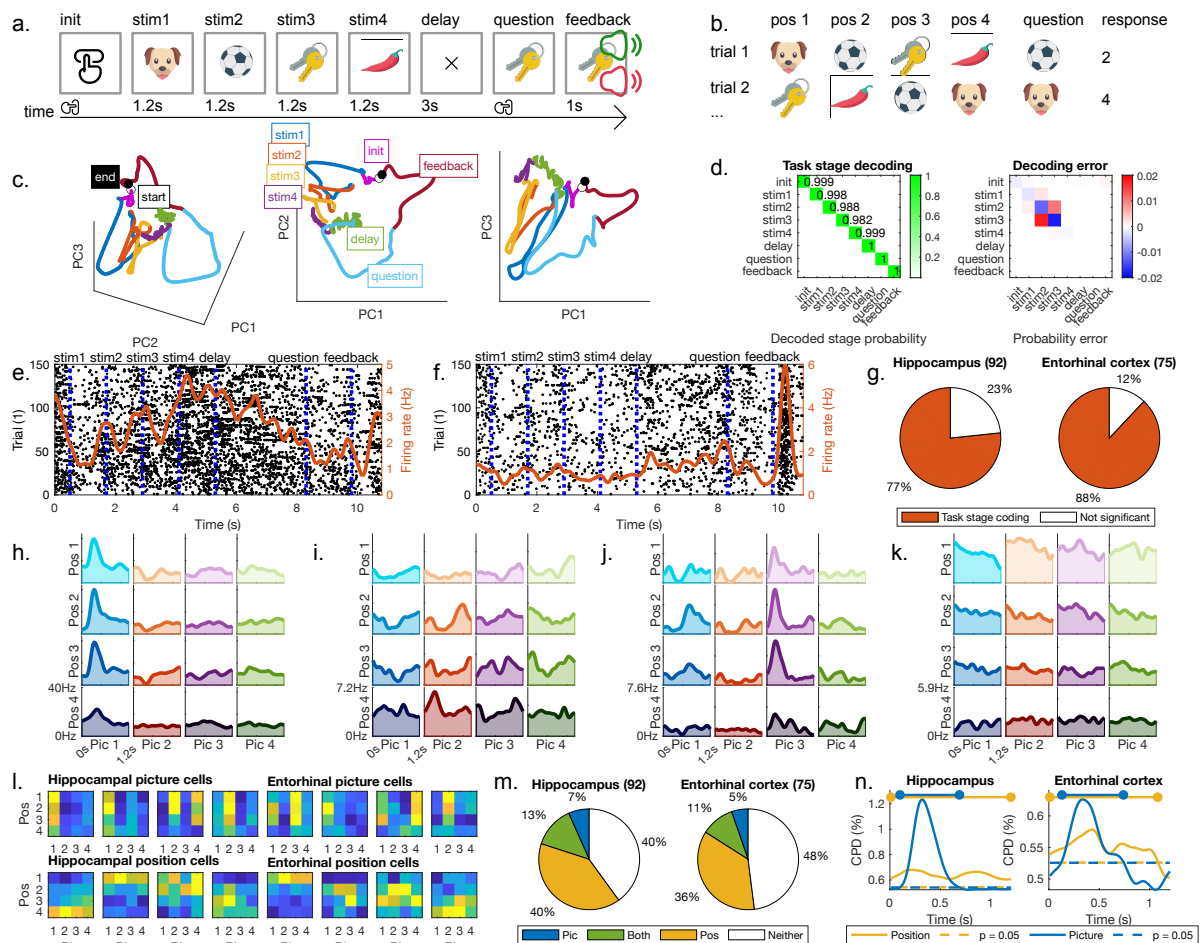


Fig 1. Single unit data. **a.** Participants had to learn the order of 4 pictures presented one after another. After a 3-second delay, we queried the sequence position of one of the pictures. **b.** On each trial, we used the same four pictures but changed the order in which they were displayed. **c.** 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: the resulting posterior probability of decoding each task stage given the firing rates of each task stage. Right: the decoding error. **e.** Raster plot across trials of a cell coding for the middle of the trial. **f.** Raster plot across trials of a cell coding for the feedback stage of the trail. **g.** Proportion of cells significantly coding for task stage. **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-l. **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).

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 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%, see example cells in Fig 1g-l and Fig S1.1, S1.2). Interestingly, while the sensory picture code was (expectedly¹⁴) 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). 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). Such a finding would support 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^{6,15–18}.

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 axis¹⁹ (Fig 2d bottom). Because obtaining enough cells along the A-P axis in human patients is impractical, we used fMRI. 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 113-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 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).

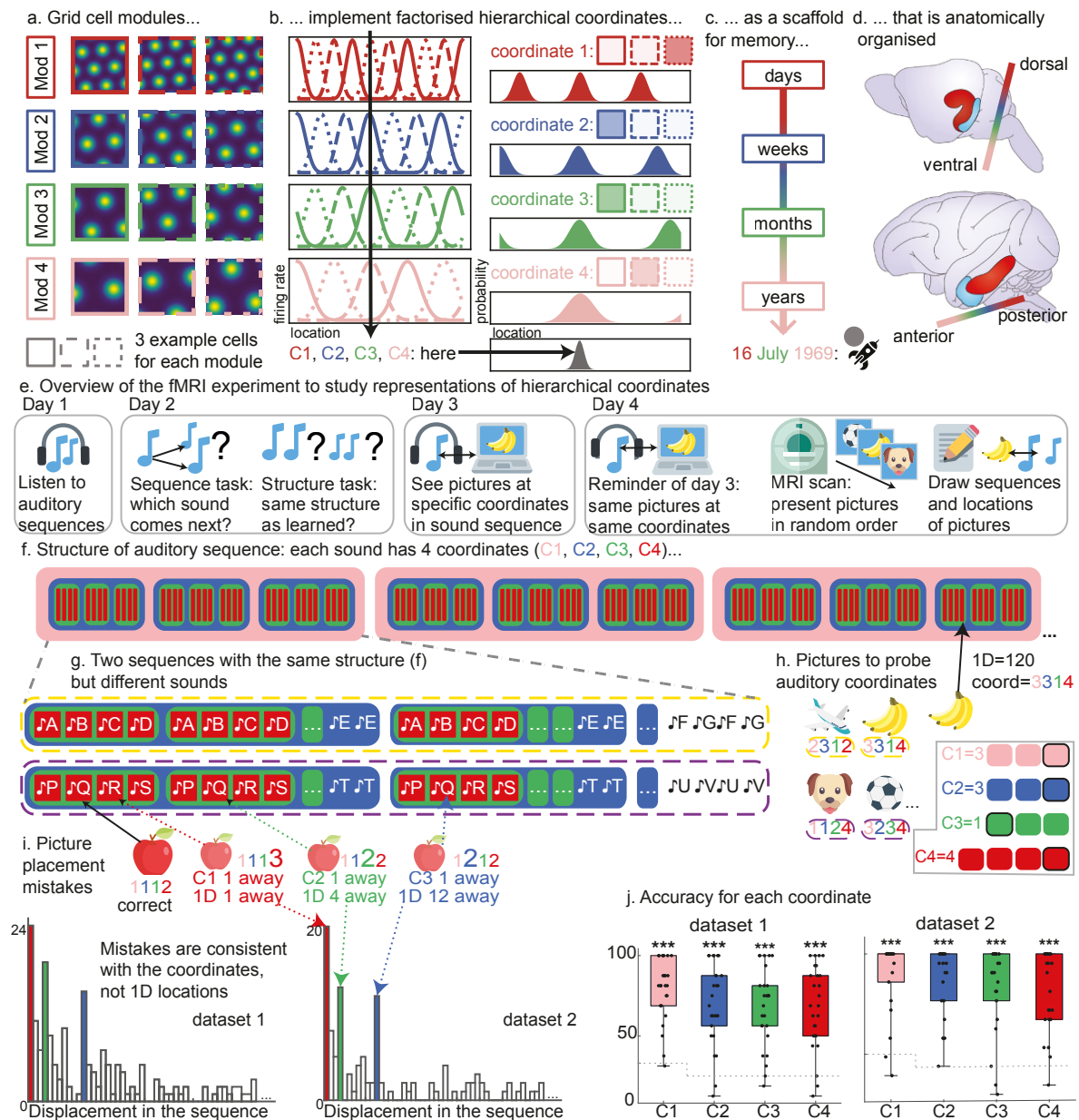


Fig 2. fMRI experimental design and behaviour. **a-d.** Schematic of logic for anatomical hypothesis: coordinates systems organised anatomically by scale in EC. **a-b.** Grid cells are a coordinate system for space. **a.** Simulated example grid rate maps. Grid cells in rodent entorhinal cortex group together in anatomically separable modules of different spatial scales (rows). Cells within a module exhibit different shifts or 'phases' (columns, in differently dashed boxes). **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.

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^{20,21} (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 16x16, 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 analysis²². The data matrix, at each voxel, was the correlation matrix between voxel responses in a local searchlight between the 16 images, resulting in a 16x16 data matrix per-voxel. We ran 4 separate regressions 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 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 orderness metric was positive for different combinations of sequence coordinates in both datasets, which confirmed our hypothesised order. First, we calculated the orderness 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 orderness 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).

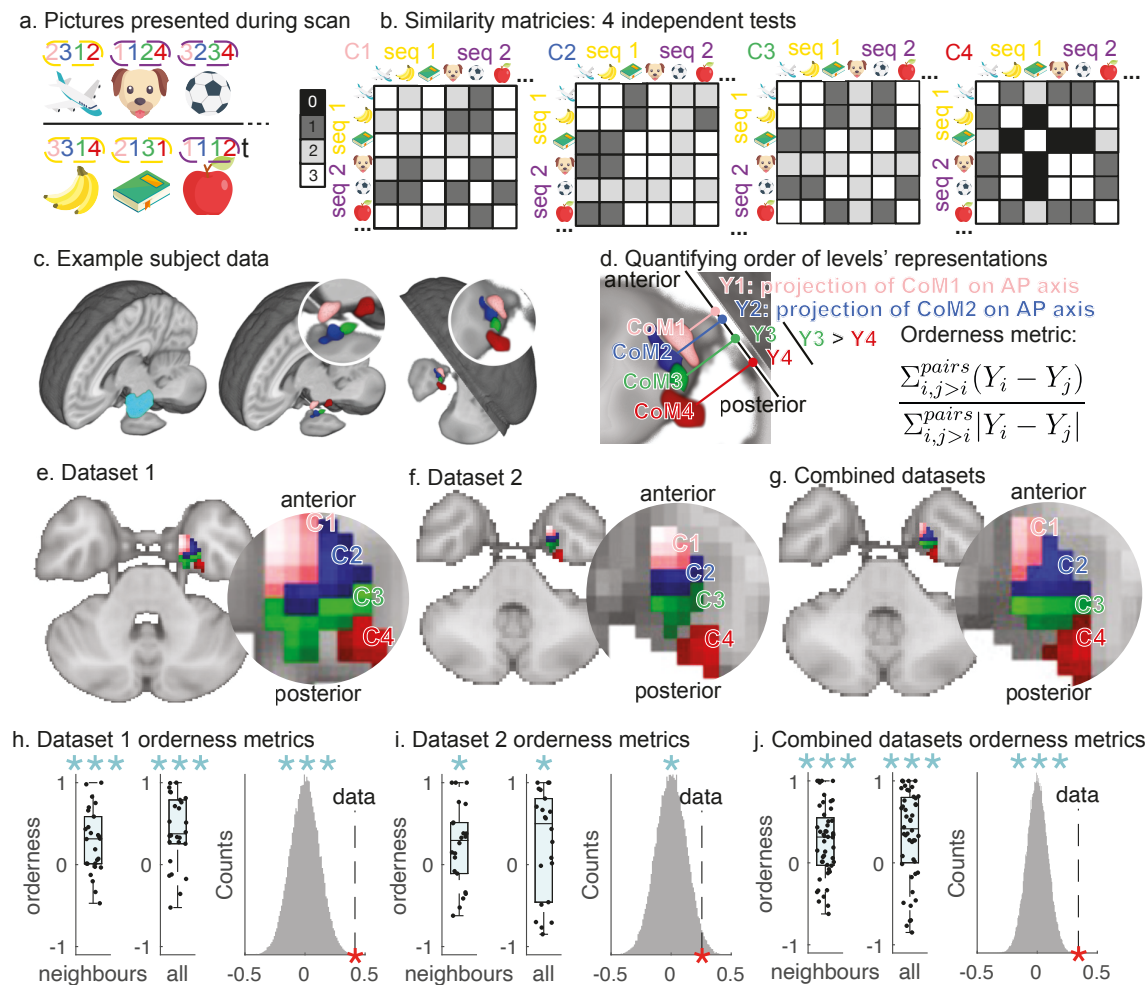


Fig 3. Coordinate representations are anatomically organised. **a.** We present pictures associated to specific coordinates in random order in the MRI scanner. **b.** To find coordinate representations, we specify hypothesised similarities between pictures with respect to each level of the hierarchy. 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 (CoM $_i$) 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).

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. 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. For completeness, we also tested the complementary analysis, using only the within-sequence similarity matrices elements. Here only the permutation test for the combined dataset was significant (Fig S4.1 combined $p < 0.05$, Supplementary Table 7).

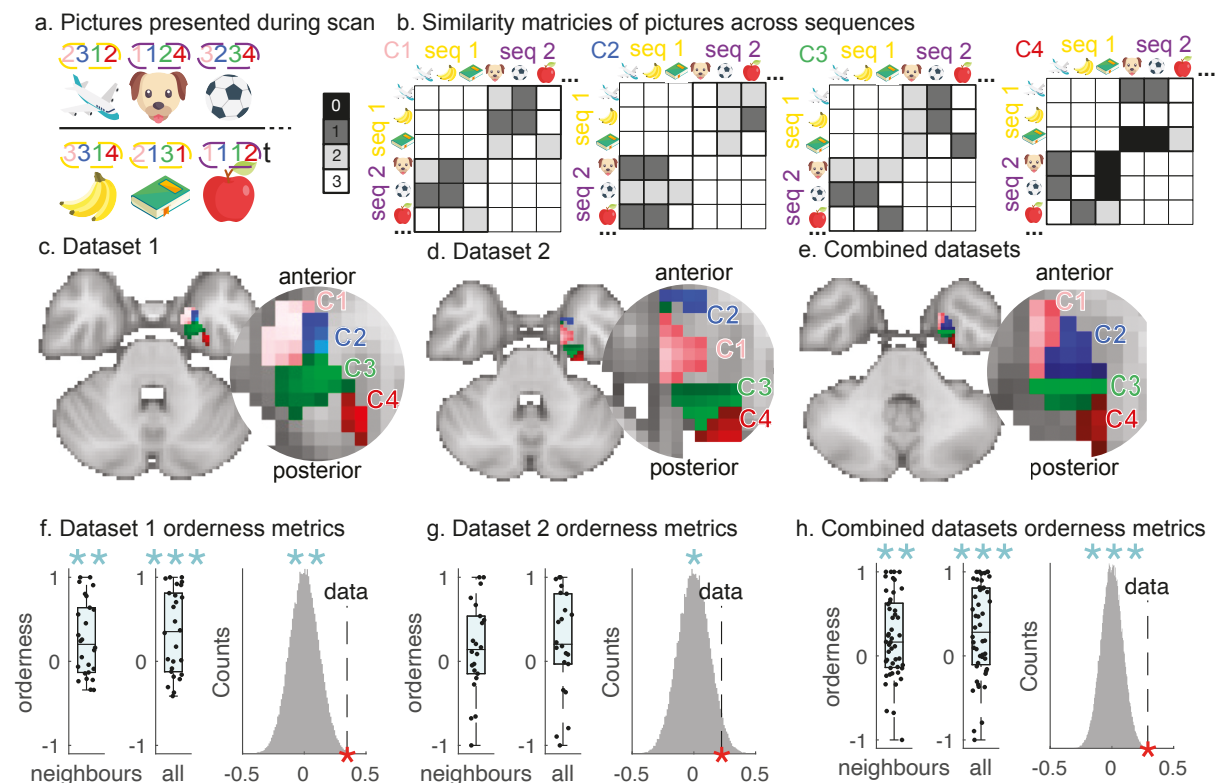


Fig 4. Abstract generalisable coordinate system. **a.** We presented pictures associated to specific coordinates in random order in the MRI scanner. Notice that some were paired with a sequence 1 coordinate during training (yellow dashed line) and others with sequence 2 (purple dashed line). **b.** Across-sequence similarity matrices. To find coordinate representations, we specify hypothesized 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**).

Discussion

Understanding the temporal structure of experiences is a key aspect of learning and memory. Here, we claim that the brain encodes this structure with an abstract and efficient coordinate system which shares representational principles across domains. Using single-unit data, we first show that in a task with a repeating structure the MTL encodes a stimuli-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.

Coordinate systems for space and time are well-established in the entorhinal cortex. Grid cells famously abstract over different spatial environments⁵ while time cells function as a clock independent of sensory inputs^{23–25}. But our experiences are not distributed equally in space and time. They are richly structured. Previous work has shown that these statistical dependencies in experience do indeed influence both fMRI signals and cellular responses in the human entorhinal cortex^{10,26–28}, and has suggested a coordinate-like representation for simple spatial sequences²⁹. Our study shows that general non-spatial sequences are also abstracted to a coordinate system and that this coordinate system efficiently compresses the experiential structure, at least when it is hierarchical.

Strikingly, this hierarchical compression aligns with the anatomical gradient of the entorhinal cortex in exactly the fashion predicted by the grid cell hierarchy. 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, and recent theoretical studies give reasons to favour both possibilities. Forming a bespoke coordinate system that captures the experiential statistics allows efficient inferences and predictions about unseen events^{6,15,30} and efficient memory search³¹. 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 limits¹³. 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 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. In this work, we only focused on the recordings from entorhinal cortex and hippocampus. 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 ^{32,33}. We used Wave clus for spike sorting and manually reviewed the outcome to identify single-unit activity versus noise³⁴.

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 across 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 posterior 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

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 73 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 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 we projected the pictures on 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 2x2mm with no breaks and a slice thickness of 2mm each. The slices were acquired in interleaved order, with an echo time (TE) of 20 s 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 x 2 x 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 = 1x1 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 (FSL³⁵). 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 MCFLIRT³⁶. Brain extraction was performed using the automated brain extraction tool BET³⁷. 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 (FLIRT³⁶ and FNIRT³⁸), 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 (RSA²⁰), 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 toolbox²¹. 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 correlations³⁹. As a result, we end up with one 16x16 data RDM, containing dissimilarities between each pair of the 16 pictures. As the last step in the within-participant RSA pipeline, we regressed the upper triangle of the data RDM separately 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 included as multiple regressors in a single GLM; 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)⁴⁰, we calculate statistical maps at the group level to assess the reliability of these results across the entire participant population⁴¹. 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 atlas anatomical mask thresholded at 50% (results for a stricter 75% mask threshold are reported in Fig S3.3, Supplementary Table 3)⁴². 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 'orderliness metric'.

$$\frac{\sum_{i,j>i}^{pairs} (Y_i - Y_j)}{\sum_{i,j>i}^{pairs} |Y_i - Y_j|}$$

The orderness metric was defined as follows: $\frac{\sum_{i,j>i}^{pairs} (Y_i - Y_j)}{\sum_{i,j>i}^{pairs} |Y_i - Y_j|}$. 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 Y₁ Y₂ Y₃ Y₄ in A-P order, and -1 opposite ordering, so Y₄ Y₃ Y₂ Y₁ 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 4x4 matrices together to get a 4x4 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 orderliness 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.

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).

Author contribution

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

References

1. Hafting, T., Fyhn, M., Molden, S., Moser, M.-B. & Moser, E. I. Microstructure of a spatial map in the entorhinal cortex. *Nature* **436**, 801–806 (2005).
2. Stensola, H. *et al.* The entorhinal grid map is discretized. *Nature* **492**, 72–78 (2012).
3. Stemmler, M. B., Mathis, A. & Herz, A. V. M. Connecting multiple spatial scales to decode the population activity of grid cells. *Sci. Adv.* **1**, e1500816 (2015).
4. Bush, D., Barry, C., Manson, D. & Burgess, N. Using Grid Cells for Navigation. *Neuron* **87**, 507–520 (2015).
5. Fyhn, M., Hafting, T., Treves, A., Moser, M.-B. & Moser, E. I. Hippocampal remapping and grid realignment in entorhinal cortex. *Nature* **446**, 190–194 (2007).
6. Whittington, J. C. R. *et al.* The Tolman-Eichenbaum Machine: Unifying Space and Relational Memory through Generalization in the Hippocampal Formation. *Cell* **183**, 1249–1263.e23 (2020).
7. Behrens, T. E. J. *et al.* What Is a Cognitive Map? Organizing Knowledge for Flexible Behavior. *Neuron* **100**, 490–509 (2018).
8. Constantinescu, A. O., O'Reilly, J. X. & Behrens, T. E. J. Organizing conceptual knowledge in humans with a gridlike code. *Science (80-.).* **352**, 1464–1468 (2016).
9. Bao, X. *et al.* Grid-like Neural Representations Support Olfactory Navigation of a Two-Dimensional Odor Space. *Neuron* **102**, 1066–1075.e5 (2019).
10. Park, S. A., Miller, D. S. & Boorman, E. D. Inferences on a multidimensional social hierarchy use a grid-like code. *Nat. Neurosci.* 2021 249 **24**, 1292–1301 (2021).
11. Bongioanni, A. *et al.* Activation and disruption of a neural mechanism for novel choice in monkeys. *Nat.* 2021 5917849 **591**, 270–274 (2021).
12. Veselic, S. *et al.* A cognitive map for value-guided choice in ventromedial prefrontal cortex. *bioRxiv* 2023.12.15.571895 (2023). doi:10.1101/2023.12.15.571895
13. Chandra, S., Sharma, S., Chaudhuri, R. & Fiete, I. High-capacity flexible hippocampal associative and episodic memory enabled by prestructured “spatial” representations. *bioRxiv* 2023.11.28.568960 (2023). doi:10.1101/2023.11.28.568960
14. Quiroga, R. Q., Reddy, L., Kreiman, G., Koch, C. & Fried, I. Invariant visual representation by single neurons in the human brain. *Nature* **435**, 1102–7 (2005).
15. Dordek, Y., Soudry, D., Meir, R. & Derdikman, D. Extracting grid cell characteristics from place cell inputs using non-negative principal component analysis. *Elife* **5**, 1–36 (2016).
16. Stachenfeld, K. L., Botvinick, M. M. & Gershman, S. J. The hippocampus as a predictive map. *Nat. Neurosci.* **20**, 1643–1653 (2017).
17. Yoon, K., Lewallen, S., Kinkhabwala, A. A., Tank, D. W. & Fiete, I. R. Grid Cell Responses in 1D Environments Assessed as Slices through a 2D Lattice. *Neuron* **89**, 1086–1099 (2016).
18. Mark, S. *et al.* Flexible neural representations of abstract structural knowledge in the human Entorhinal Cortex. *Elife* **13**, (2024).
19. Strange, B. A., Witter, M. P., Lein, E. S. & Moser, E. I. Functional organization

- of the hippocampal longitudinal axis. *Nat. Rev. Neurosci.* **15**, 655–669 (2014).
20. Kriegeskorte, N., Mur, M. & Bandettini, P. A. Representational similarity analysis - connecting the branches of systems neuroscience. *Front. Syst. Neurosci.* **2**, 4 (2008).
21. Nili, H. *et al.* A Toolbox for Representational Similarity Analysis. *PLoS Comput. Biol.* **10**, e1003553 (2014).
22. Kriegeskorte, N., Goebel, R. & Bandettini, P. A. Information-based functional brain mapping. *Proc. Natl. Acad. Sci.* **103**, 3863–3868 (2006).
23. Kraus, B. J., Robinson, R. J., White, J. A., Eichenbaum, H. & Hasselmo, M. E. Hippocampal ‘Time Cells’: Time versus Path Integration. *Neuron* **78**, 1090–1101 (2013).
24. Allen, T. A., Salz, D. M., McKenzie, S. & Fortin, N. J. Nonspatial sequence coding in CA1 neurons. *J. Neurosci.* **36**, 1547–1563 (2016).
25. Umbach, G. *et al.* Time cells in the human hippocampus and entorhinal cortex support episodic memory. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 28463–28474 (2020).
26. Garvert, M. M., Moutoussis, M., Kurth-Nelson, Z., Behrens, T. E. J. & Dolan, R. J. Learning-induced plasticity in medial prefrontal cortex predicts preference malleability. *Neuron* **85**, 418–428 (2015).
27. Tacikowski, P., Kalender, G., Ciliberti, D. & Fried, I. Human hippocampal and entorhinal neurons encode the temporal structure of experience. *Nat.* 2024 1–8 (2024). doi:10.1038/s41586-024-07973-1
28. Baram, A. B., Muller, T. H., Nili, H., Garvert, M. M. & Behrens, T. E. J. Entorhinal and ventromedial prefrontal cortices abstract and generalize the structure of reinforcement learning problems. *Neuron* **109**, 713–723.e7 (2021).
29. Tsitsiklis, M. *et al.* Single-Neuron Representations of Spatial Targets in Humans. *Curr. Biol.* **30**, 245–253.e4 (2020).
30. Sreenivasan, S. & Fiete, I. Grid cells generate an analog error-correcting code for singularly precise neural computation. *Nat. Neurosci.* 2011 1410 **14**, 1330–1337 (2011).
31. McNamee, D. C., Stachenfeld, K. L., Botvinick, M. M. & Gershman, S. J. Flexible modulation of sequence generation in the entorhinal–hippocampal system. *Nat. Neurosci.* 2021 246 **24**, 851–862 (2021).
32. Boran, E. *et al.* Persistent hippocampal neural firing and hippocampal-cortical coupling predict verbal working memory load. *Sci. Adv.* **5**, (2019).
33. Boran, E., Hilfiker, P., Stieglitz, L., Sarnthein, J. & Klaver, P. Persistent neuronal firing in the medial temporal lobe supports performance and workload of visual working memory in humans. *Neuroimage* **254**, (2022).
34. Chaure, F. J., Rey, H. G. & Quiñ Quiroga, R. A novel and fully automatic spike-sorting implementation with variable number of features. *J. Neurophysiol.* **120**, 1859–1871 (2018).
35. Jenkinson, M., Beckmann, C. F., Behrens, T. E. J., Woolrich, M. W. & Smith, S. M. FSL. *Neuroimage* **62**, 782–790 (2012).
36. Jenkinson, M., Bannister, P., Brady, M. & Smith, S. Improved Optimization for the Robust and Accurate Linear Registration and Motion Correction of Brain Images. *Neuroimage* **17**, 825–841 (2002).
37. Smith, S. M. Fast robust automated brain extraction. *Hum. Brain Mapp.* **17**, 143–155 (2002).
38. Jenkinson, M. & Smith, S. A global optimisation method for robust affine registration of brain images. *Med. Image Anal.* **5**, 143–156 (2001).

39. Cai, M. B., Schuck, N. W., Pillow, J. W. & Niv, Y. Representational structure or task structure? Bias in neural representational similarity analysis and a Bayesian method for reducing bias. *PLOS Comput. Biol.* **15**, e1006299 (2019).
40. Woolrich, M. W., Ripley, B. D., Brady, M. & Smith, S. M. Temporal Autocorrelation in Univariate Linear Modeling of FMRI Data. *Neuroimage* **14**, 1370–1386 (2001).
41. Woolrich, M. W., Behrens, T. E. J., Beckmann, C. F., Jenkinson, M. & Smith, S. M. Multilevel linear modelling for FMRI group analysis using Bayesian inference. *Neuroimage* **21**, 1732–1747 (2004).
42. Amunts, K., Mohlberg, H., Bludau, S. & Zilles, K. Julich-Brain: A 3D probabilistic atlas of the human brain's cytoarchitecture. *Science (80-.).* **369**, 988–992 (2020).

Supplementary Material



Figure S1.1. 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 (rows) in a particular position (column), then plot the 4x4 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).

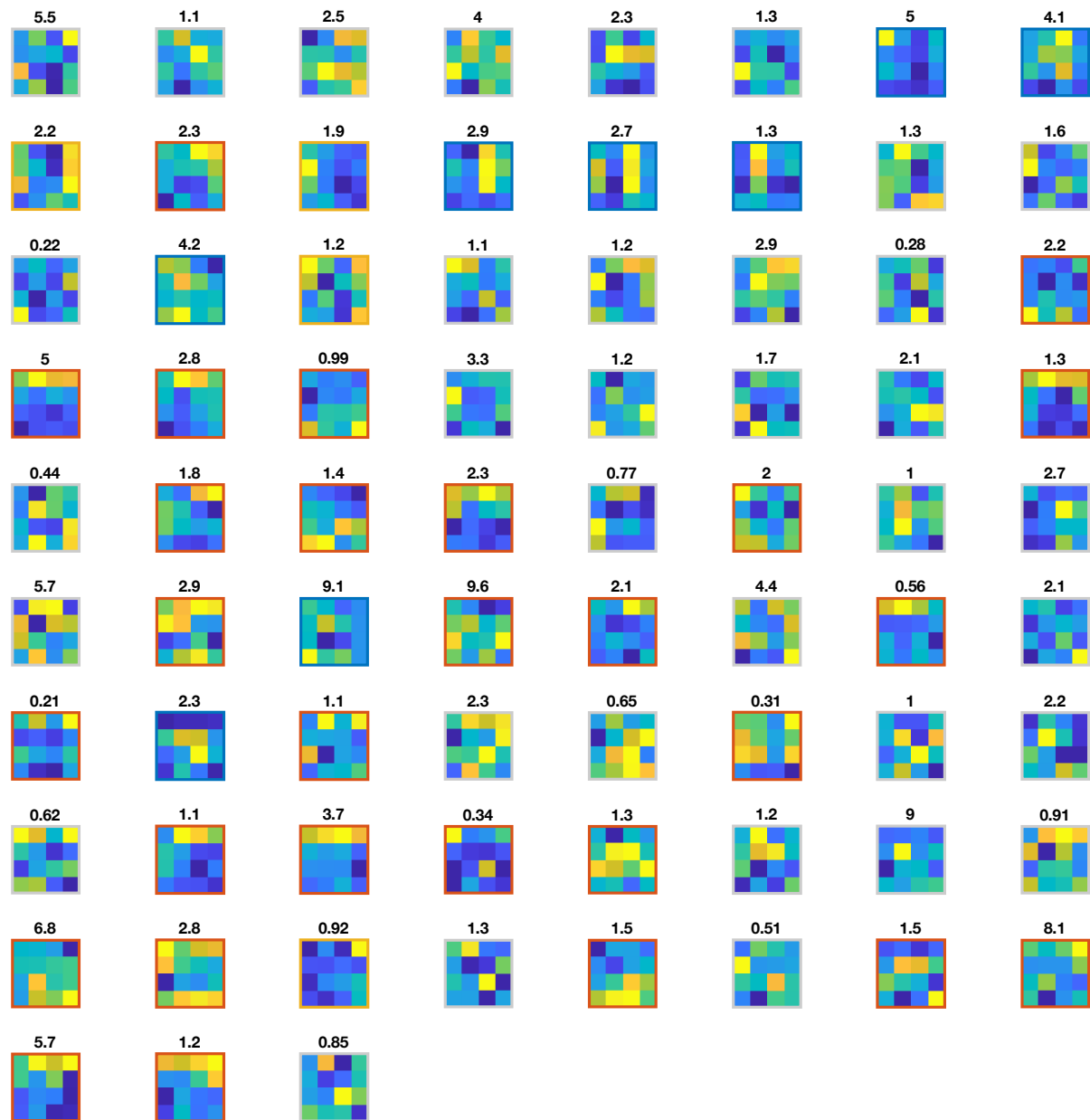


Figure S1.2. Summary firing rate plots for all entorhinal cells. For each cell we average firing rates across time and across trial stages with a particular picture (rows) in a particular position (column), then plot the 4x4 firing rate map with the maximum firing rate in yellow and the minimum firing 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).

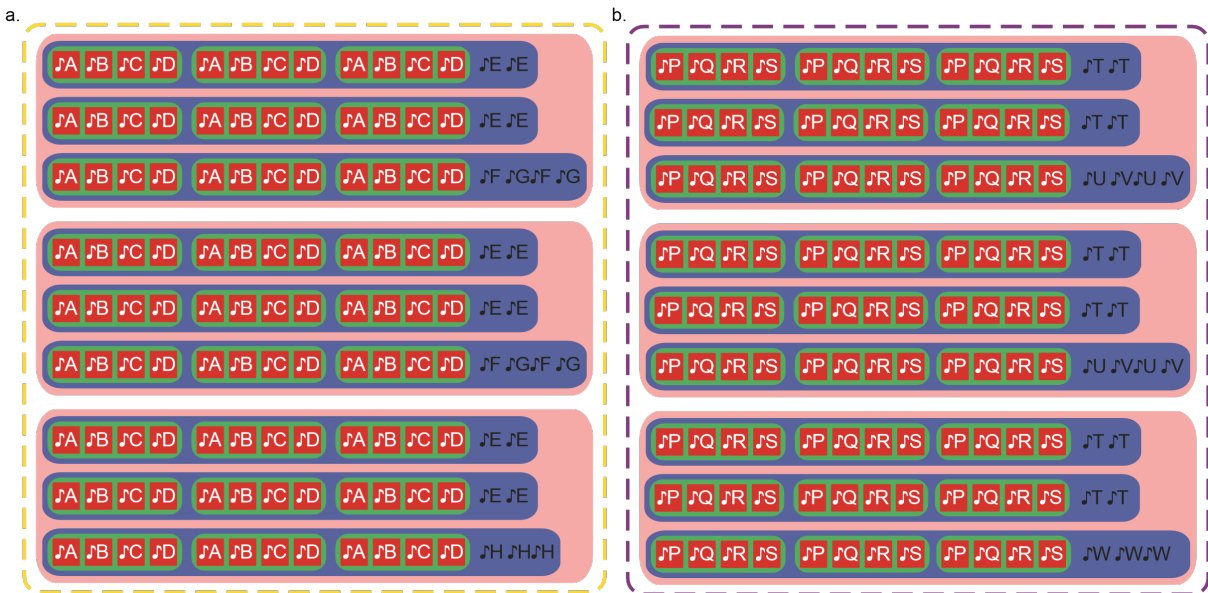
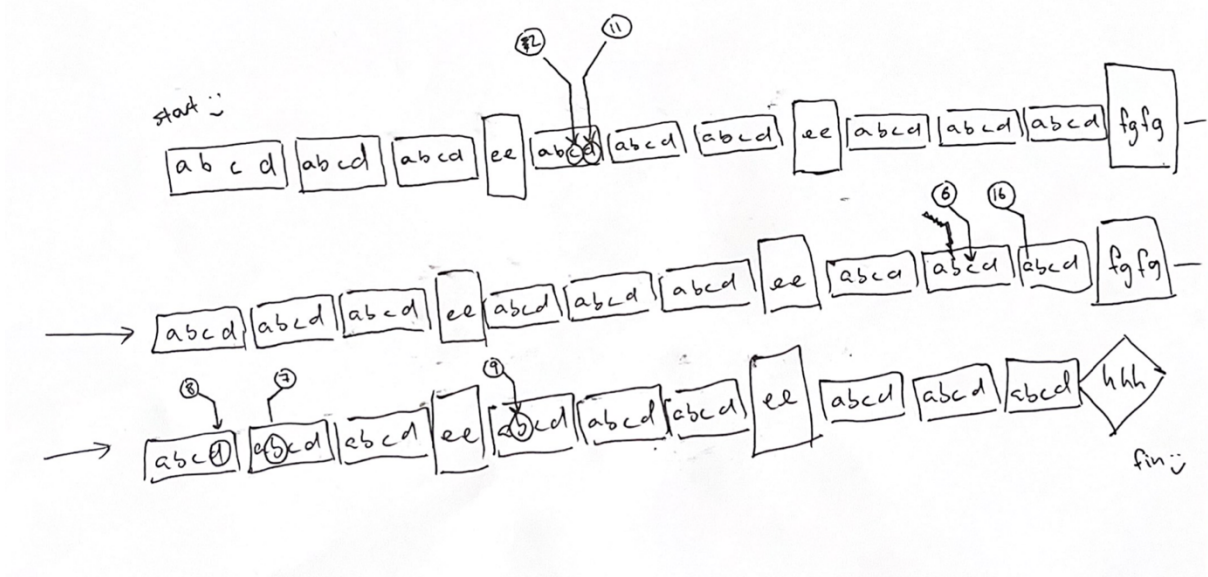
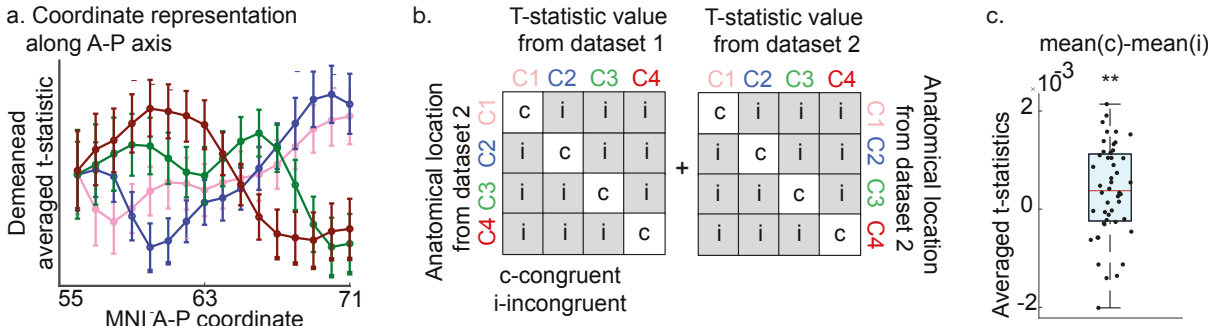


Fig S2.1. Auditory sequences. Participants learnt two sequences with identical structures but different sounds.



Supplementary Fig 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.

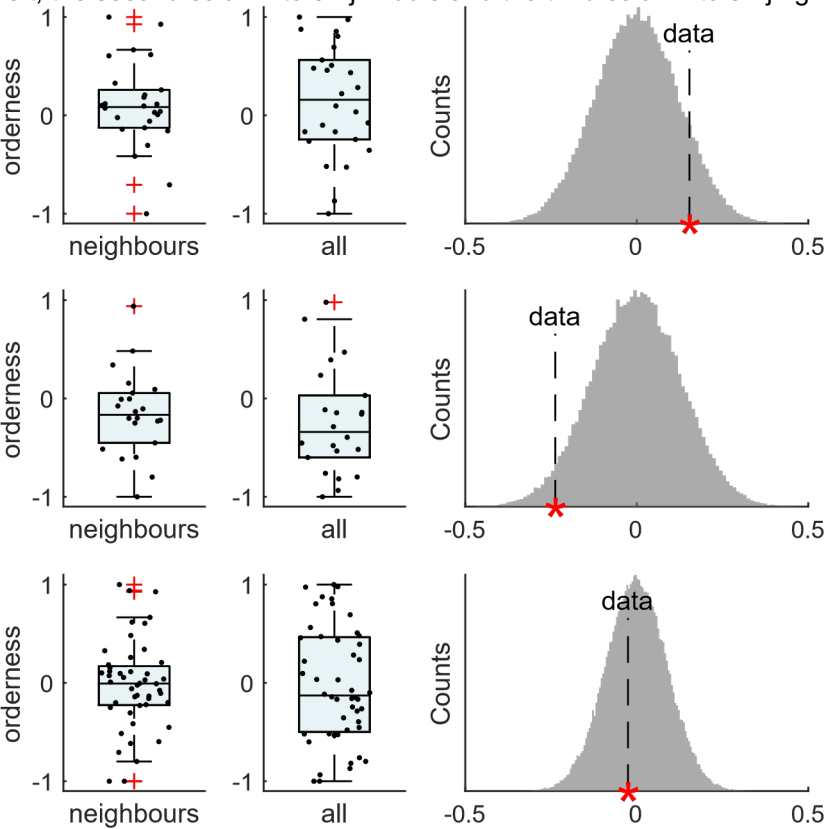


Supplementary figure S3.1. a. T-statistic for each level's coordinate representation (C1 is pink, C2 is blue, C3 is green, C4 is red) averaged across participants within A-P slices as a function of slice A-P position in MNI space. Note that the peaks are ordered according to the hypothesised order (both

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, $p=0.0039$.

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

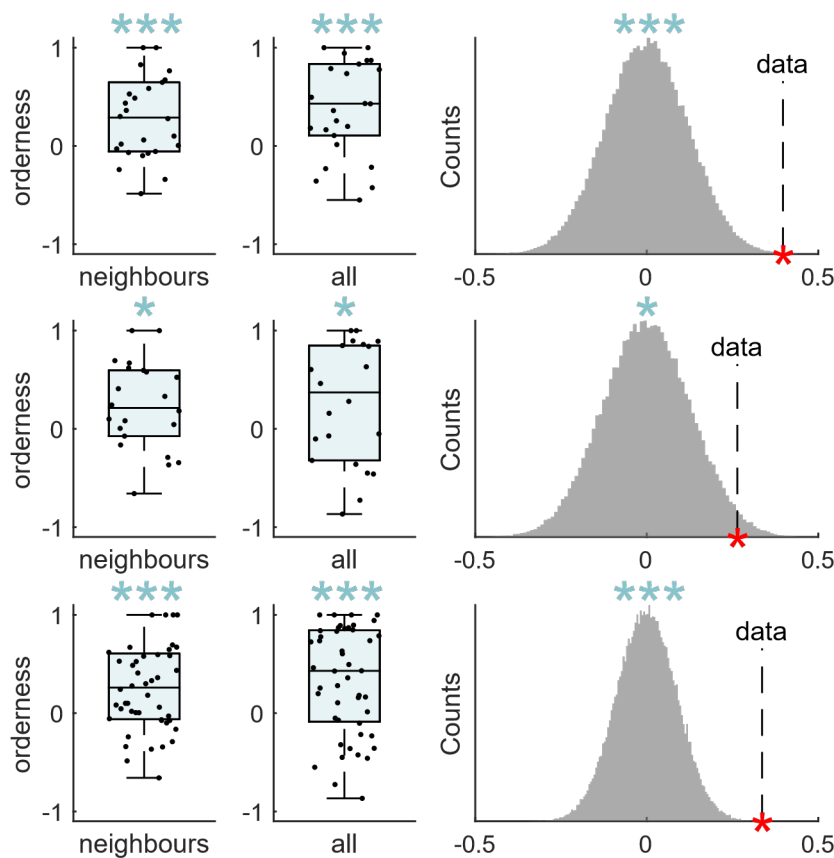
Supplementary Table 1. P-values table for Fig 3. P-values for orderness metric for neighbouring pairs of levels (left column), including all pairs of levels (middle column), and averaged across participants against a null distribution from shuffled level labels (right column) in dataset 1 (first row), 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.



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.

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

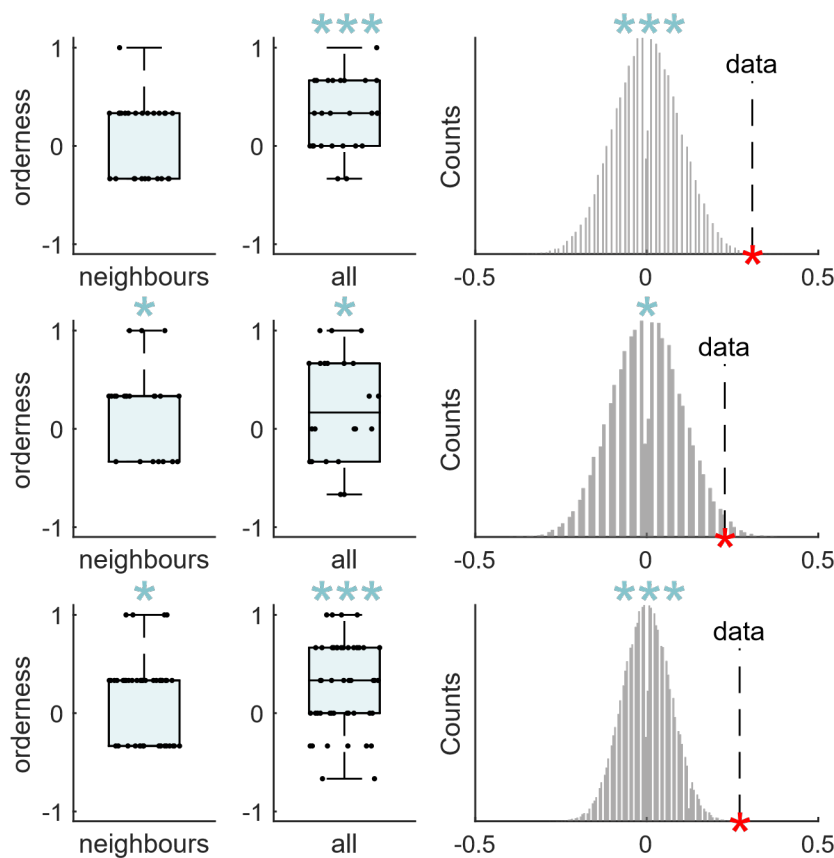
Supplementary Table 2. P-values table, corresponding to the Fig S3.2, for the orderness metrics computed on centres of mass defined in the right entorhinal mask.



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.

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

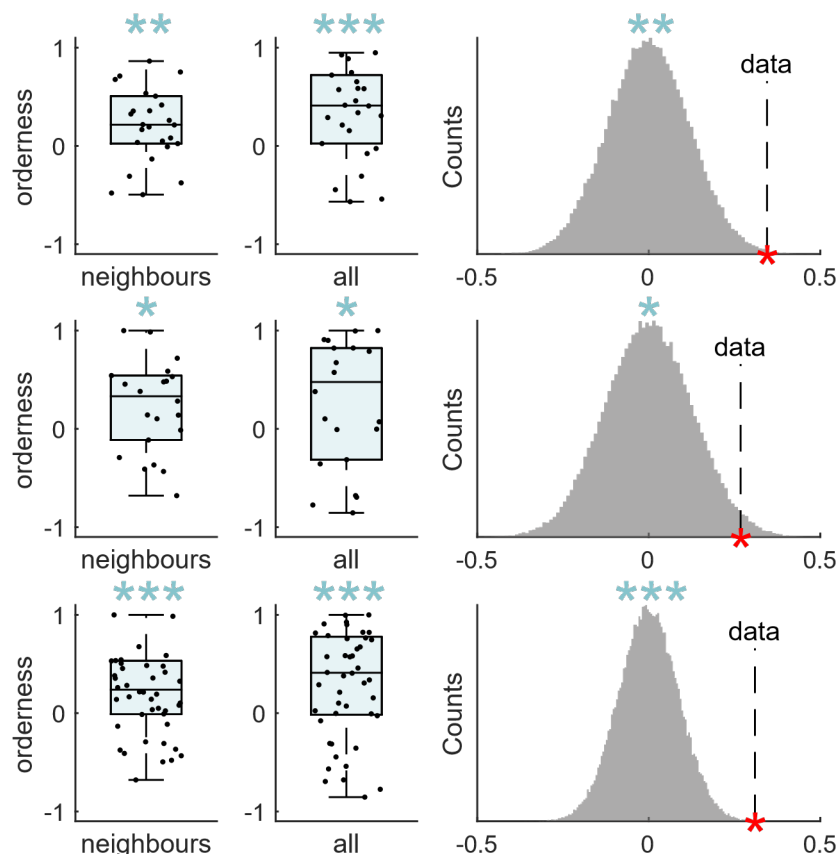
Supplementary Table 3. P-values table, corresponding to the Fig S3.3, for the orderness metrics computed on centres of mass in the stricter entorhinal mask.



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.

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



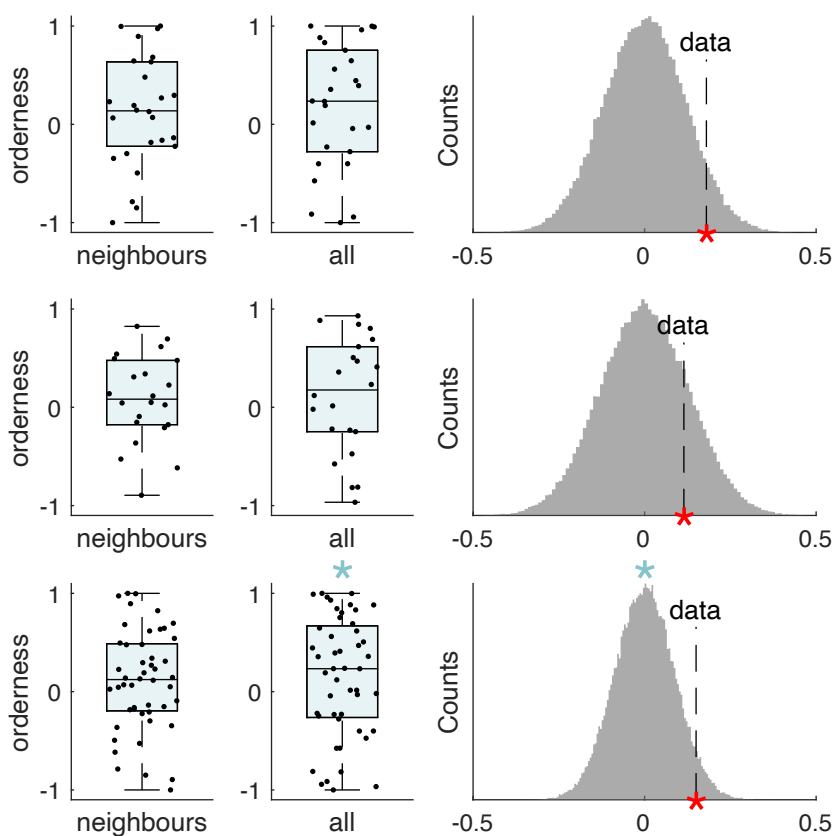
Supplementary Fig S3.5. The orderness metrics computed on centres of mass defined on statistical maps that result from a **multiple regression** (fitting 4 similarity regressors together (one for each level), Methods). Orderness metric for neighbouring pairs of levels (left column), including all pairs 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.

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

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 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.



Supplementary Fig S4.1. Within-sequence similarities matrices (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.

Data	Neighbours	All	Permutations
Dataset 1	0.13863	0.07912	0.06638
Dataset 2	0.18975	0.18788	0.19055
Both	0.0783	0.04714	0.0446

Supplementary Table 7. P-values table, corresponding to the Fig S4.1