Behavior-dependent directional tuning in the human visual-navigation network

Matthias Nau[1,2*], Tobias Navarro Schröder[1], Markus Frey[1,2], Christian F. Doeller[1,2]

[1]Kavli Institute for Systems Neuroscience, NTNU, Trondheim, Norway.
[2]Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany.
*Corresponding author. Email: matthias.nau@ntnu.no

Abstract

The brain derives cognitive maps from sensory experience to guide memory formation and behavior. Despite extensive efforts, it still remains unclear how the underlying population activity relates to active behavior and memory performance. Here, we combined 7T-fMRI with a kernel-based encoding model of virtual navigation to map world-centered directional tuning across the human cortex. First, we present an in-depth analysis of directional tuning in visual, retrosplenial and parahippocampal cortices and the hippocampus. Second, we show that tuning strength, width and topology of this directional code during memory-guided navigation depend on successful encoding of the environment. Finally, we show that participants' locomotory state influences this tuning in sensory and mnemonic regions such as the hippocampus. We demonstrate a direct link between neural population tuning and human cognition and show that high-level memory processing interacts with network-wide environmental coding in the service of behavior.

Introduction

Human scene processing and navigation regions interface between the lower-level sensory and higher-level cognitive domain. They gradually construct world-centered mnemonic representations of the environment, a cognitive mapping process thought to culminate in the medial temporal lobe (MTL) (Byrne et al., 2007; Bicanski and Burgess, 2018; Epstein and Baker, 2019; Epstein et al., 2017; Clark et al., 2018; Vann et al., 2009; Nau et al., 2018a). Areas such as the retrosplenial cortex (Vann et al., 2009) and the parahippocampal cortex represent the spatial layout (Epstein and Kanwisher, 1998) and 3D structure (Lescroart and Gallant, 2019) of the currently viewed scene, as well as its relative openness (Kravitz et al., 2011) or its boundaries (Ferrara and Park, 2016). Downstream regions like the entorhinal cortex and the hippocampus use this information to derive a stable representation of the world and one's own position, direction and speed in it (Hartley et al., 2014; Moser et al., 2017). Together, such spatial representations are often referred to as 'cognitive map' and are thought to fundamentally shape our memories and guide behavior (O'Keefe and Nadel, 1978). To understand this process, we believe it is critical to study the neural population activity of these regions in a naturalistic setting and in the light of the behavior they support.

A critical challenge the brain needs to solve to map the environment is keeping track of our own direction as we move. Previous studies revealed directional representations and activity related to heading perception in several areas including the medial parietal lobe and retrosplenial cortex (Baumann and Mattingley, 2010; Chadwick et al., 2015; Marchette et al., 2014; Shine et al., 2016), the parahippocampal gyrus (Bellmund et al., 2016; Doeller et al., 2010; Kim and Maguire, 2019; Marchette et al., 2014; Vass and Epstein, 2013, 2016), the entorhinal/subicular region (Chadwick et al., 2015; Doeller et al., 2010; Jacobs et al., 2010; Shine et al., 2019; Vass and Epstein, 2016), the thalamus (Shine et al., 2016) and the superior parietal cortex (Gourtzelidis et al., 2005; Schindler and Bartels, 2013). Also, hippocampal and parahippocampal activity encodes heading in the horizontal plane (Indovina et al., 2013). Most of these studies used dedicated and constrained directional judgment- and mental imagery tasks and often examined direction in a self-centered frame of reference. To date, it remains unclear how cognitive mapping is mediated by the scene processing and navigation network and how active spatial behavior and memory relate to environmental processing in this pathway.

Here, we used 7T functional magnetic resonance imaging (fMRI) to monitor human brain activity during naturalistic virtual navigation in a spatial memory task (Fig. 1). Inspired by prior successes of encoding models in characterizing fMRI responses in other domains (Kriegeskorte and Douglas, 2019; Naselaris et al., 2011), we then developed an iterative kernel-based encoding model (Fig. 2) of the participants' navigation behavior (Supplementary Fig. 1) to map directional tuning across the human cortex. Importantly, our model treats direction in a world-centered frame of reference. We

further analyzed the impact of spatial memory performance and locomotory states on this tuning, focusing on scene processing and navigation regions due to their proposed involvement in cognitive mapping and known directional coding (Bicanski and Burgess, 2018; Epstein and Baker, 2019; Epstein et al., 2017). These regions include the early visual and retrosplenial cortex, the parahippocampal gyrus, the entorhinal cortex as well as the hippocampus.

Our objectives were twofold. First, we aimed to quantify and map directional tuning in the human scene processing and navigation network during active spatial behavior. Second, we examined how this tuning relates to the participant's behavior and memory.

Results

During fMRI scanning, participants freely navigated in a circular virtual reality (VR) arena via key presses while memorizing and reporting object locations within it (Fig. 1A). Across different trials, participants indicated the locations of these hidden objects by navigating to them. After each trial, they received feedback about the true object location before the next trial started. We then tracked the improvement in memory performance over trials by assessing the 'memory error', i.e. the Euclidean distance between true and remembered location in each trial (Fig. 1B, Supplementary Fig. 2).

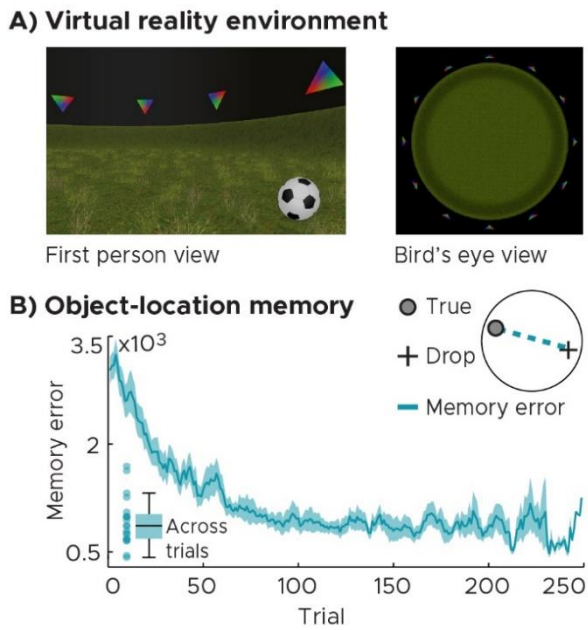


Figure 1. Spatial memory task in virtual reality (VR). A) First person and bird's eye view of the VR-environment in which participants navigated freely via key presses. The circular arena was surrounded by 12 landmarks matched in visual features (colored triangles). Across trials, participants memorized and reported object locations by navigating to them and pressing a 'drop' button followed by feedback. B) Object-location memory. Participants' memory performance improved as indicated by a decrease in memory error (Euclidean distance between drop and true location). The blue line and shaded area represent the mean and SEM of the memory error across participants. Data were smoothed with a moving average kernel of 5 trials. The inset depicts the median memory error across trials for single participants and as whisker-boxplot (center, median; box, 25th to 75th percentiles; whiskers, 1.5 × interquartile range).

Our here developed encoding model analysis comprises multiple individual steps. We modelled world-centered virtual head direction (vHD) using basis sets of circular-gaussian vHD-kernels (Fig. 2A). Next, we estimated voxel-wise weights for each kernel with ridge regression using a training data set (Fig. 2B, Supplementary Fig. 3A). We then used these weights to predict the time course of

each voxel in an independent test set (Fig. 2C). We define directional tuning strength as the model performance, i.e. how well the model predicted the time course of a voxel or region. Finally, by iterating through multiple basis sets differing in the number and the full-width-at-half-maximum (FWHM) of the directional kernels (Supplementary Fig. 3B) we also estimated the corresponding tuning width of each voxel. Tuning width was defined as the FWHM of the kernels leading to the best model performance. Finally, we examined the relationship between the tuning strength and width estimated for different regions, the participants' navigation behavior itself as well as their performance in the spatial memory task.

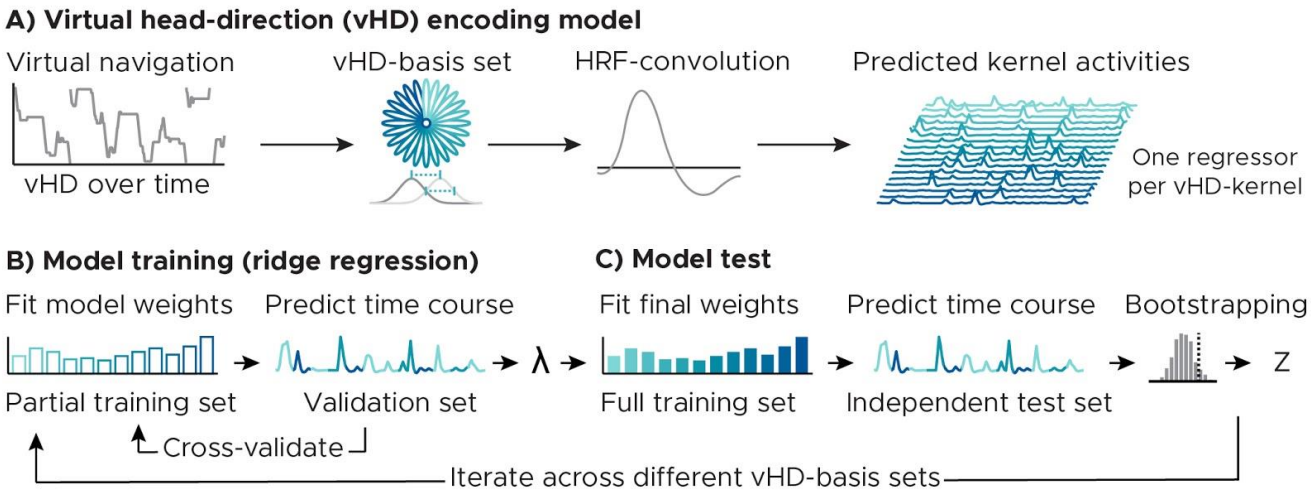


Figure 2. Analysis logic. A) Virtual head direction (vHD) encoding model. We modeled vHD using multiple basis sets of circular-gaussian kernels covering the full 360°. Given the observed vHD, we generated predicted time courses (regressors) for all kernels in each basis set. The basis sets differed in the (full-width-at-half-maximum) kernel width and number. Spacing and width were always matched to avoid over-representing certain directions. The resulting regressors were convolved with the hemodynamic response function (SPM12) to link the kernel activity over time to the fMRI-signal. B) Model training. We estimated voxel-wise weights for each regressor in a training data set (80% of all data of a participant or 4 runs, 10 minutes each) using ridge regression. To estimate the L2-regularization parameter (λ), we again split the training set into partial training (60% data, 3 runs) and validation sets (20% data, one run). Weights were estimated in the partial training set, and then used to predict the time course of the validation set via Pearson correlation. This was repeated for ten values of λ (log-spaced between 1-10.000.000) and cross-validated such that each training partition served as validation set once. We then used the λ that resulted in the highest average Pearson's R to fit the final model weights using the full training set. C) Model test: We used the final model weights to predict each voxel's time course in an independent test set (held-out 20% data, always the run halfway through the experiment) via Pearson correlation. These Pearson correlations were used to test model performance on a voxel-by-voxel level (Fig. 3,4). For a regions of interest analysis (Fig. 5, 6), we additionally converted model performance into Z-scores via bootstrapping, ensuring that the results reflected effects of kernel width and not of number. The null distribution of each voxel was obtained by weight-shuffling (k = 500). Both model training and test were repeated for all basis sets.

Our results are presented in three sections. First, we establish how the vHD-encoding model works by mapping directional tuning strength and width across the cortex. Second, we demonstrate that the strength, width and topology of this tuning depend on the participants' spatial memory performance. On a behavioral level, we show evidence that this likely relates to how well the environment has been encoded. Finally, we show that the tuning in both sensory and high-level mnemonic regions reflected the behavioral state of our participant's, i.e. whether they were moving

or not. Notably, our results cannot be explained by biases in sampling (Supplementary Fig. 1), model regularization or data quality (Supplementary Fig. 3).

Mapping directional tuning during spatial navigation

Participants navigated in a VR-environment, memorizing and reporting object locations within it. We used an iterative kernel-based voxel-wise encoding model of vHD to map directional tuning strength and width across the cortex (Fig. 2). The model performance is the Pearson correlation between the voxel time course predicted by the model and the one observed in the test set. The model prediction builds on weights estimated for each kernel using an independent model training procedure.

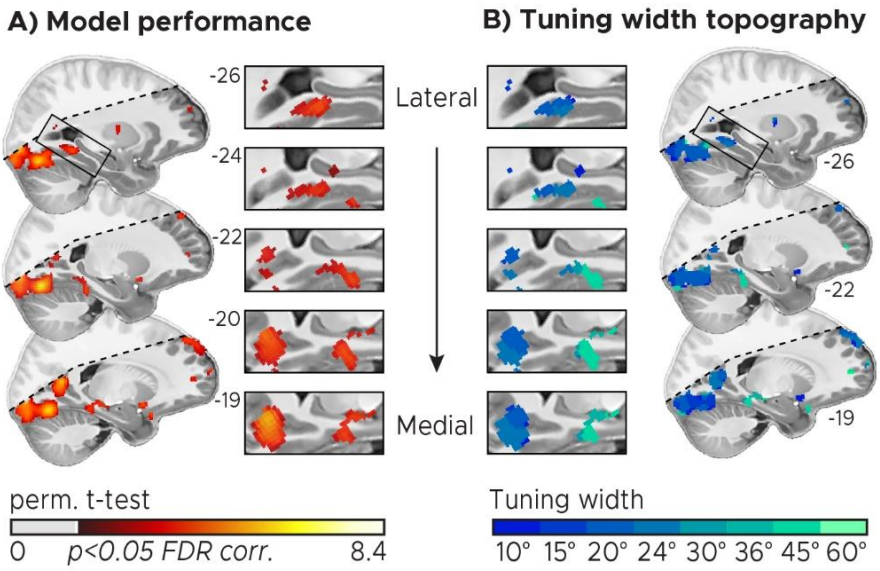


Figure 3. Mapping directional tuning across the human cortex (n = 20). A) Tuning map: directionally tuned voxels were determined by testing model performance against zero on group-level using a permutation-based one-sample t-test (Nichols and Holmes, 2002) *with k = 10000 shuffles. We plot pseudo-T-maps thresholded at p<0.05, FDR-corrected for all basis sets at T1-resolution overlaid on the group average T1-scan. Approximate MNI coordinates were added. Inserts zoom in on the parahippocampal cortex and the medial parietal cortex. Multiple regions in the occipital lobe, the medial parietal and temporal lobes and the parahippocampal gyrus were directionally tuned. B) Tuning width: For each directionally tuned voxel, we color coded the median tuning width (full-width-at-half-maximum of the directional kernels in the optimal basis set) that led to the highest pseudo-T-value (depicted in A). The tuning width follows a narrow-to-broad topology along the parahippocampal long-axis. Shaded regions fell outside the scanning field of view in at least one participant.*

Our model successfully predicted activity in multiple regions in the ventral occipital and medial parietal cortex as well as in the MTL (Fig. 3A). These regions overlap with known scene processing and navigation regions such as the retrosplenial cortex (Vann et al., 2009) and the parahippocampal cortex as well as with the posterior hippocampal formation (Epstein and Baker, 2019; Epstein et al., 2017). Along the parahippocampal long-axis, the tuning width followed a narrow-to-broad topology: narrow kernels best predicted activity in more posterior parts, wider kernels in anterior parts of the left-hemispheric parahippocampus (Fig. 3B).

Directional tuning reflects spatial memory performance

After establishing that our vHD-encoding model did indeed predict activity in the visual-navigation network, we next asked whether the tuning was related to successful encoding of the environment and the object locations in it. We hypothesized that the tuning should be stronger in ventral visual stream and medial temporal regions in participants that performed well in the spatial memory task. This hypothesis built on the idea that stronger tuning should indicate enhanced retrieval of directional information from high-level mnemonic systems.

To test this, we repeated the group level analysis depicted in Fig. 3A, this time splitting the participants into two groups based on their across-trial median memory error (median split, Fig. 1D). We found that our model predicted activity in strikingly different networks in these two participant groups (Fig. 4A), the direction of these effects however was opposite of what we had predicted. In participants with low memory error, i.e. good memory performance, the model predicted activity in the medial and ventral occipital lobe. Strikingly, in participants with high memory error, the model predicted activity in the parahippocampal gyrus and in the MTL. A group-level permutation-based rank-correlation between memory error and model performance further indicated that these differences build on a systematic relationship between directional tuning and spatial memory (Supplementary Fig. 4). In both groups, we observed bilateral clusters in the medial parietal lobe. These clusters however barely overlapped between groups (Fig. 4B). In the low-memory-error group the model predicted activity in a more anterior part, in the high-memory-error group in a more posterior part of the medial parietal lobe, akin to previous reports of an anterior-posterior functional distinction in this region (Silson et al., 2016, 2019).

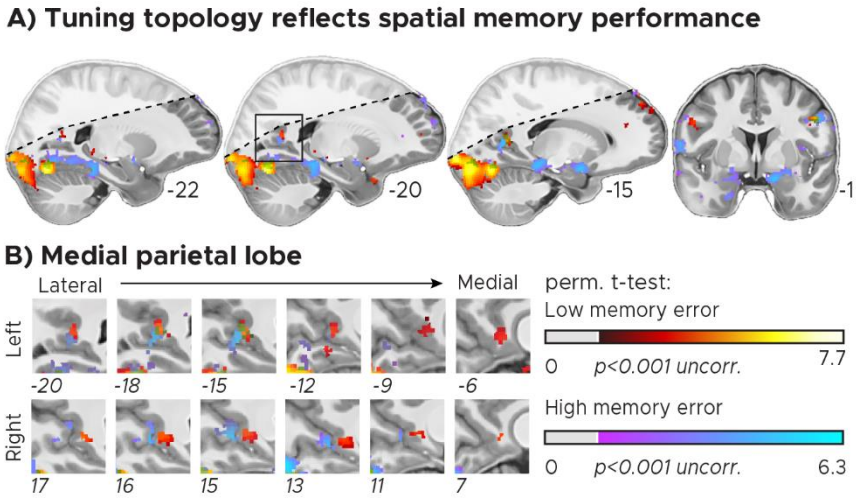


Figure 4. Directional tuning topology reflects spatial memory performance. Participants were split into two groups depending on their across-trial median memory error (2 x n = 10). A) Tuning maps: directionally tuned voxels tuning were determined by testing model performance against zero on group-level using a permutation-based t-test (Nichols and Holmes, 2002) *with 1024 unique random possible shuffles. This procedure results in a minimal possible p-value of 0.00098, precluding FDR-*

correction. We therefore plot pseudo-T-maps thresholded at p<0.001 uncorrected for all basis sets at T1-resolution overlaid on the group-average T1-scan. Hot colors depict results for the low-memory-error group, cool colors for the high-memory-error group. Approximate MNI-coordinates added. B) Zoomed in depiction of the medial parietal lobe/retrosplenial cortex (RSC). There is an anterior-posterior distinction in directional tuning in RSC as a function of spatial memory performance. Shaded regions fell outside the scanning field of view in at least one participant.

To further characterize directional tuning explicitly in regions that derive world-centered representations of the visual environment, we next conducted a regions of interest (ROI) analysis focusing on the early visual cortex (EVC), the retrosplenial cortex (RSC), parahippocampal gyrus (PHG), the posteromedial entorhinal cortex (pmEC) and the hippocampus (HPC) (Fig. 5A). We tested the pmEC subdivision of the entorhinal cortex because its rodent homologue region (Maass et al., 2015; Navarro Schröder et al., 2015) is known to encode direction (Giocomo et al., 2014; Kornienko et al., 2018; Sargolini et al., 2006). We obtained the vHD-model performance for every voxel in our ROIs and every directional basis set as described before. For the ROI analysis, we added a bootstrapping procedure to ensure that our results reflected an effect of kernel width, and not of kernel number. For every basis set, we converted the Pearson correlations into Z-scores via weight-shuffling (Fig. 2C) and then averaged across the 25% most reliable training voxels in each ROI to reduce noise (see methods for shuffling and voxel selection details). The resulting Z-scores expressed how well the model predicted the activity relative to the voxel's null distribution. For each participant group and ROI, we selected the basis set that led to the best model performance on average (Fig. 5B). If a given region was not directionally tuned, the corresponding Z-scores should be zero. We tested this on group-level (one-tailed) as well as a difference between groups (two-tailed) using permutation-based t-tests (Fig. 5C, see methods for details). We observed that EVC and RSC encoded direction in both participant groups (EVC: low memory error: $t(9) = 2.85$, $p = 0.014$, $p^{FDR} = 0.048$; high memory error: $t(9) = 4.00$, $p = 0.004$, $p^{FDR} = 0.027$; RSC, low: $t(9) = 3.04$, $p = 0.006$, $p^{FDR} = 0.041$; high: $t(9) = 2.62$, $p = 0.015$, $p^{FDR} = 0.040$). Importantly, in PHG and pmEC such tuning was observed only in participants that had a high memory error (PHG, low: $t(9) = 0.62$, $p = 0.320$; high: $t(9) = 2.16$, $p = 0.028$, $p^{FDR} = 0.040$; pmEC, low: $t(9) = -0.42$, $p = 0.661$; high: $t(9) = 2.59$, $p = 0.020$, $p^{FDR} = 0.040$, in line with the voxel-wise group results (Fig. 4). In pmEC, the tuning strength additionally differed between groups ($t(18) = 2.32$, $p = 0.036$). To again test whether this group difference reflected a systematic relationship between tuning strength and memory error, we conducted a post-hoc permutation-based rank-correlation between memory error and pmEC model performance on ROI level, which indeed seconded these results (rho = 0.48, $p = 0.035$, $k = 10000$, Supplementary Fig. 4). Notably, while the rodent homologue of pmEC is known to encode world-centered direction (Giocomo et al., 2014; Kornienko et al., 2018; Sargolini et al., 2006), another entorhinal subregion, the anterolateral entorhinal cortex (alEC) is not. Consistently, we observed directional tuning only in pmEC, not in alEC (Supplementary Fig. 6).

In addition to the tuning strength, our approach also allowed to estimate the tuning width for each ROI. For each individual participant and ROI, we selected the tuning width that led to the optimal model performance (Supplementary Fig. 5) and compared it between groups. Strikingly, while above-mentioned model performance in RSC was matched, the tuning width differed between groups ($t(18) = 2.04$, $p = 0.044$). Participants with high memory error had a sharper tuning in RSC than participants with low memory error (Fig. 5D).

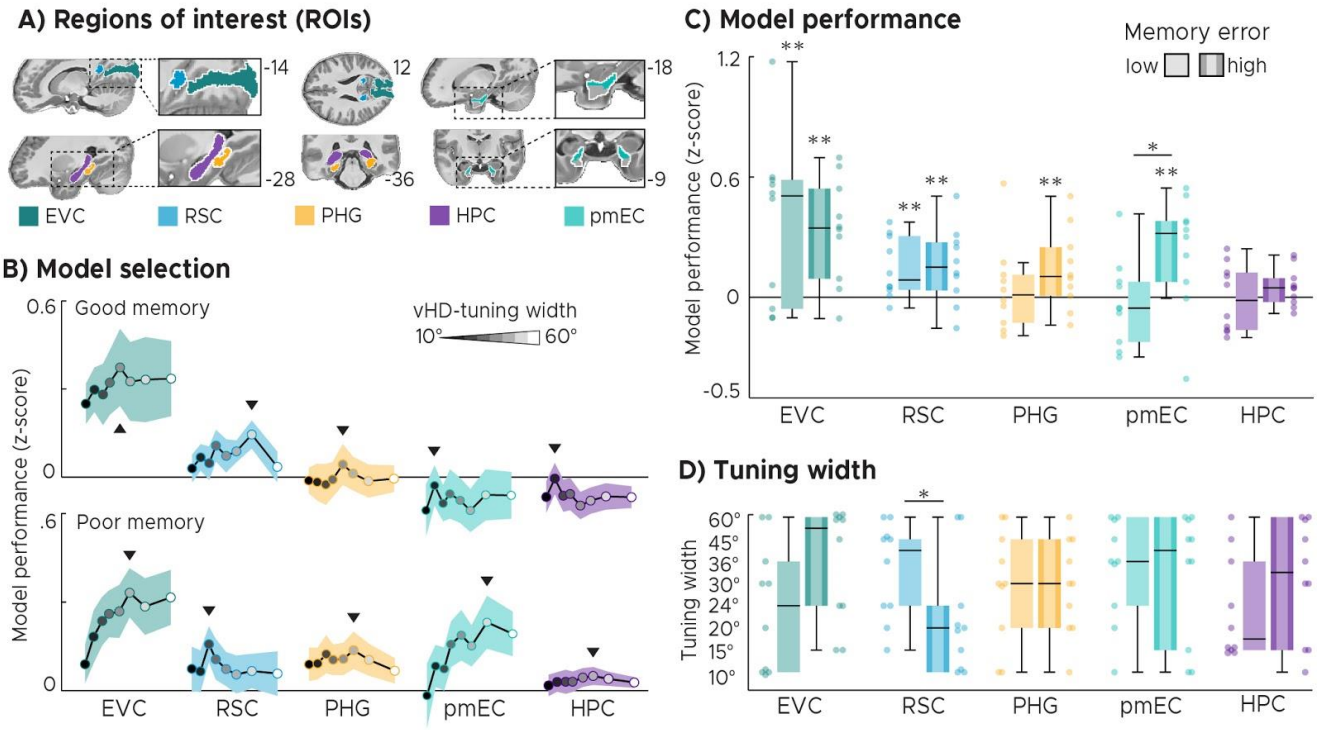


Figure 5. Regions of interest (ROI) analysis. A) Human scene processing and navigation regions tested in this study: Early visual cortex (EVC), Retrosplenial cortex (RSC), Parahippocampal gyrus (PHG), posteromedial entorhinal cortex (pmEC) and the hippocampus (HPC). B) Model selection: We plot the model performance (Z-score) on ROI-level for all basis sets. The black line and the shaded area represent the mean and SEM across participants. Each dot represents the group-average model performance for one basis set, with darker colors representing narrow kernels and lighter colors representing wider kernels. The following kernel widths were tested: 10°, 15°, 20°, 24°, 30°, 36°, 45°, 60°. The black triangles mark the basis set that lead to the optimal model performance. C) Optimal model performance for the two (high- and low-memory-error) participant groups. We plot single participant data and group-level whisker-boxplots (center, median; box, 25th to 75th percentiles; whiskers, 1.5 × interquartile range). We observed directional tuning in EVC and RSC in both groups. In PHG and pmEC this tuning depended on spatial memory performance. D) Optimal tuning width. Similar to B,C we plot the tuning width that led to the highest memory performance selected on individual participant level. Participants with low memory error had wider tuning than the ones with high memory error. Hence, unlike tuning strength, the tuning width in RSC reflected spatial memory performance.

In sum, while EVC, RSC, PHG and pmEC were directionally tuned in at least one of the participant groups, tuning strength in pmEC as well as tuning width in RSC strikingly reflected how well participants performed in the spatial memory task. This is in line with our hypothesis that the tuning should indicate whether the environment has been successfully encoded or not. However, we had hypothesized that stronger directional tuning in higher-level visual and MTL regions should be

associated with better spatial memory performance. Our empirical test suggested the opposite. Why would mnemonic regions be more directionally tuned in participants that performed poorly in the spatial memory task? One explanation could be the following. Contrary to our initial hypothesis, the differences between groups could reflect the participants' ongoing effort in encoding rather than retrieving a map of the environment. If so, this predicted that participants with high median memory error were still in the process of mapping the environment and the object locations in it. In turn, they should be improving more in the spatial memory task than participants with an already low memory error. To test this post-hoc hypothesis, we analyzed the trial by trial memory error in more detail (Fig. 1B, Supplementary Fig. 2A). Indeed, we found that while both participant groups approached the same level of memory performance, the low-memory-error participants reached ceiling earlier than the high-memory-error participants (Supplementary Fig. 2A,B). We also estimated the learning rate for each participant, i.e. the slope of a linear regression line fit to the memory error across trials. The high-memory-error participants had steeper slopes ($t(18) = 2.65$, $p = 0.019$, $k = 10000$, Supplementary Fig. 2C), again suggesting stronger improvements in memory performance compared to the low-memory-error participants.

Directional tuning reflects the behavioral state

To investigate the relationship between directional tuning and the participants' behavior further, we repeated above-described ROI-analysis twice, once only modeling periods in which the participants moved, and once in which they stood still. In both cases, the participants rotated (Supplementary Fig. 1A,B). We compared these two scenarios by contrasting the respective model performances (Fig. 6A), revealing a positive effect of locomotion on tuning strength in EVC and RSC (EVC: locomotion: $t(19) = 4.04$, $p = 0.0004$, $p^{FDR} = 0.003$; stationary: $t(19) = 1.38$, $p = 0.091$; contrast: $t(19) = 2.06$, $p = 0.049$; RSC: locomotion: $t(19) = 3.68$, $p = 0.001$, $p^{FDR} = 0.004$; stationary: $t(19) = -0.82$, $p = 0.79$; contrast: $t(19) = 2.20$, $p = 0.043$). Conversely, the activity in PHG and HPC could be better predicted while participants stood still (PHG: locomotion: $t(19) = -0.91$, $p = 0.799$; stationary: $t(19) = 2.37$, $p = 0.015$; contrast: $t(19) = -2.35$, $p = 0.024$; HPC: locomotion: $t(19) = -3.48$, $p = 0.999$; stationary: $t(19) = 1.77$, $p = 0.046$; contrast: $t(19) = -3.18$, $p = 0.004$, $p^{FDR} = 0.028$, Fig. 5C). Our results suggest that the directional tuning in human scene processing and navigation regions reflects not only spatial memory performance, but also the behavioral state of the participants (i.e. whether they move or not). Notably, the fact that the differences in tuning strength and width were region-specific (Fig. 6A), suggest that our results cannot be explained by general differences in statistical power between the two behavioral states.

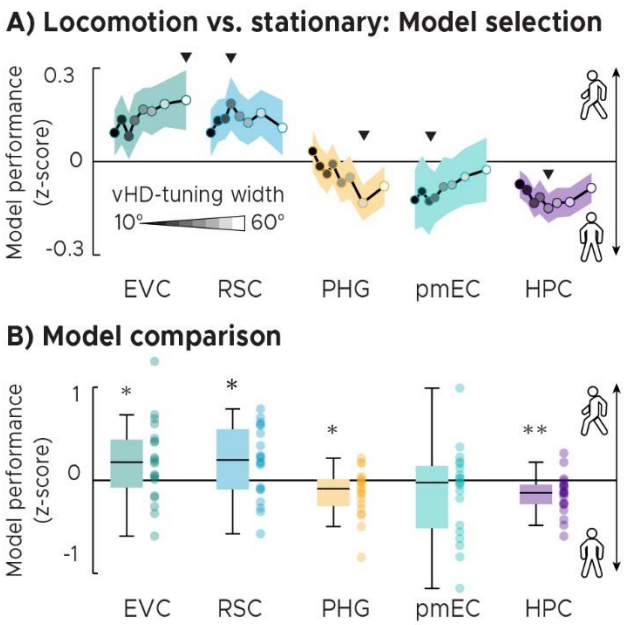


Figure 6. Behavioral state analysis. The analysis described in Fig. 2 was repeated, this time separating periods when participants navigated and when they stood still. A) Model selection. We plot the difference in model performance (Z-score) between locomotion and stationary periods across tuning widths (grayscale dots represent tuning width: narrow:dark, wide:light., also see Fig. 5). Positive values indicate that voxel time courses in an ROI could be better predicted when participants locomoted. Negative values indicate the opposite, with better model performance during stationary periods. Triangles mark the kernel width leading to the strongest difference between models. B) Model comparison. We plot the difference in model performance indicated in (A) as single participant data and group-level whisker-boxplots (center, median; box, 25th to 75th percentiles; whiskers, 1.5 × interquartile range). EVC and RSC tended to be better predicted during locomotion, PHG and HPC could be better predicted during stationary periods. These results suggest that the tuning in visual and mnemonic regions depend on the locomotory state.

Discussion

The present study investigated fMRI-proxies of neural population activity reflecting directional coding during active spatial behavior. We put a focus on human scene processing and navigation regions due to their proposed involvement in cognitive mapping: they derive world-centered mnemonic representations of the environment from sensory experiences. We used 7T-fMRI to monitor brain activity of participants navigating in a virtual environment and performing a spatial memory task. We developed an iterative kernel-based encoding model of the navigation behavior to map directional tuning across the human cortex. In addition, examine its relationship to behavior and memory in detail. Visual, retrosplenial, parahippocampal, entorhinal and hippocampal regions showed distinct response profiles, with a narrow-to-broad tuning width topology along the left-hemispheric parahippocampal long-axis. Furthermore, we examined the relationship between the tuning in each region, the participants' navigation behavior and the performance in the spatial memory task. We found that the tuning in the RSC and pmEC, and likely in the parahippocampal gyrus, reflected how accurately participants reported the location of objects in the environment. Strikingly, while it was the tuning strength in pmEC and PHG, it was the tuning width and topology in RSC that depended on memory. The strength, width and topology of directional tuning were therefore associated with the spatial memory performance of the participants, which demonstrates a direct link between neural population coding and cognition. The direction of this effect as well as behavioral evidence speaks to the idea that the underlying mechanism is related to encoding rather than retrieval of environmental information. Finally, the tuning in visual, retrosplenial and

parahippocampal regions, but especially in the hippocampus, additionally signaled the behavioral state of the participants.

Directional representations in the human brain

Our observations emphasize the central role of scene processing and navigation regions in spatial cognition (Bicanski and Burgess, 2018; Byrne et al., 2007; Clark et al., 2018; Epstein and Baker, 2019; Epstein et al., 2017; Mitchell et al., 2018; Nau et al., 2018a; Vann et al., 2009) and are consistent with previous work on directional representations in the human brain (Baumann and Mattingley, 2010; Bellmund et al., 2016; Chadwick et al., 2015; Doeller et al., 2010; Indovina et al., 2013; Jacobs et al., 2010; Kim and Maguire, 2019; Marchette et al., 2014; Shine et al., 2016, 2019; Vass and Epstein, 2013). They are also consistent with lesion studies showing that damage to regions like the RSC can impair the ability to orient oneself relative to landmarks (Aguirre, 1999). Importantly, the present study goes beyond previous reports in several aspects. Most studies used dedicated and constrained directional judgment- and imagery tasks to reveal directional representations in the brain. Here, we examined directional tuning during active naturalistic navigation (in VR) and additionally demonstrate that it depends on behavioral factors and on memory. Many of these studies also examined self-centered directional coding, whereas our approach examines vHD explicitly in a world-centered frame of reference. Furthermore, using an iterative encoding model we were also able to extract additional parameters such as the tuning width from fMRI responses. Similar approaches have been used for example to map the retinotopic (Dumoulin and Wandell, 2007; Wandell and Winawer, 2015) and semantic organization of the cortex (Huth et al., 2016; Mitchell et al., 2008), the 3D-depth tuning of scene processing regions (Lescroart and Gallant, 2019), to identify perceived stimuli from brain activity (Kay et al., 2008; Thirion et al., 2006) and to reconstruct the online content of working memory (Rademaker et al., 2019; Sprague et al., 2016). Unlike previous work, the encoding model developed here (Fig. 2) does not build on information about a stimulus but importantly is informed by the behavior of our participants directly.

Visual information and head direction

An open question is whether the tuning observed here reflects processing of visual information or head direction (HD). Neurons representing world-centered HD, or HD-cells, are abundant in the brain and have been studied most intensively in rodents (Cullen and Taube, 2017), but also in monkeys (Robertson et al., 1999). HD-cells reference facing direction relative to known landmarks

(Taube and Burton, 1995) and are often compared to an ‘internal compass’ mediating our sense of direction (Butler et al., 2017; Cullen and Taube, 2017). This HD-cell compass plays a central role in cognitive mapping and is thought to mediate homing, reorientation and path integration behavior (Butler et al., 2017; Cullen and Taube, 2017; Dudchenko et al., 2019; van der Meer et al., 2010; Valerio and Taube, 2012; Weiss and Derdikman, 2018; Weiss et al., 2017). We observed that multiple brain areas encoded direction, many of which overlap with regions known to contain HD-cells in rodents and monkey. These regions include the RSC (Chen et al., 1994; Cho and Sharp, 2001; Jacob et al., 2017), the postsubiculum (part of the hippocampal formation) (Taube et al., 1990) and the entorhinal cortex (EC) (Giocomo et al., 2014; Kornienko et al., 2018; Sargolini et al., 2006). The latter consists of at least two subdivisions in rodents, the medial (MEC) and the lateral entorhinal cortex (LEC), likely corresponding to the pmEC and alEC in humans (Maass et al., 2015; Navarro Schröder et al., 2015). HD-cells have been observed only in the MEC, not the LEC, paralleling our observations of world-centered directional tuning in the human pmEC, not in the alEC (Supplementary Fig. 6).

We also observed directional tuning in the early visual cortex for which no HD-cells have been reported to date. This raises the question whether the effects reported here are due to lower-level sensory processing, not HD? We believe these two options are not exclusive and the underlying processes might strongly interact in our naturalistic task. Locomotion (Ayaz et al., 2013; Saleem et al., 2013) and world-centered location (Fournier et al., 2019; Saleem et al., 2018) have for example been shown to modulate EVC activity in rodents. Here, we observed that locomotion had a positive effect on model performance in EVC and RSC, suggesting stronger directional tuning in these regions during running (Fig. 6). Also, in monkeys (Galletti et al., 1984) and humans (Nau et al., 2018b), the EVC can represent the velocity of motion in a world-centered frame of reference, likely due to feedback from higher-level areas. Critically, high-level mnemonic regions such as the entorhinal cortex are not known to be visually responsive, yet we still observed a directional code there. Consistent with this are reports of HD-cells in MEC, the rodent homologue of the pmEC (Giocomo et al., 2014; Kornienko et al., 2018; Sargolini et al., 2006) as well as its well-connected position within the HD-circuit (Witter et al., 2017). Also, locomotion had a positive effect on model performance in EVC and a negative effect in the hippocampus, again contrasting higher-level MTL function against early visual processing. Areas such as the PHG code direction even in the absence of visual experience (Bellmund et al., 2016). In addition, shared information and activity covariations between the EVC and the hippocampal formation suggest that early perceptual processing could well be modulated by higher-level cognitive processes (Bosch et al., 2014; Fournier et al., 2019; Hindy et al., 2016; Kok and Turk-Browne, 2018; Meister and Buffalo, 2016; Nau et al., 2018a; Saleem et al., 2018).

Interestingly, the fact that the directional tuning we observed depends on behavior is at odds with the activity profile of ‘classical’ HD-cells. First, these cells provide a directional representation that is continuous (i.e. they always maintain similar activity on population-level following attractor-dynamics (Cullen and Taube, 2017; Dudchenko et al., 2019; Skaggs et al., 1995; Taube, 2007). Second, impairments or even lesions of the HD-cell system only have very moderate effects on behavior (Dudchenko et al., 2019). However, growing evidence suggests that there are at least two types of HD-cells in the brain: ‘classical’ continuous HD-cells, as well as non-continuous sensory HD-cells (Dudchenko et al., 2019). The latter have been shown to switch between active and inactive states, providing a directional representation that is controlled by visual landmarks instead of attractor-dynamics (Kornienko et al., 2018). Our present results are in line with the function and the location of these sensory HD-cells found in the entorhinal and parahippocampal cortex (Kornienko et al., 2018) as well as potentially in the RSC (Jacob et al., 2017). Some of these cells also alternate between several preferred directions depending on context, potentially explaining the variation in RSC tuning width we observed (Fig. 5D). On a population level, directional activity referenced to multiple landmarks would likely be interpreted as having broader tuning curves than the stable unidirectional counterparts. While classical HD-cells integrate vestibular inputs (Cullen and Taube, 2017; Taube, 2007), which were not available in our VR-task, visual information alone is sufficient to drive HD-cell-like coding in VR (Acharya et al., 2016) and to determine its alignment to the environment (Coletta et al., 2018). Behavioral evidence further suggests that the visual environment, specifically its geometry, influences our sense of direction (Cheng, 1986; Julian et al., 2015). In our task, the VR-arena was circular and all landmarks were carefully matched with respect to their visual appearance (Fig. 1). In one half of the arena the triangle-shaped landmarks faced upwards, in the other half downwards (Fig. 1B). The hypothetical axis in between did not contain any landmarks and did not bias vHD-sampling (Supplementary Fig. 1).

Taken together, low-level sensory information alone can explain neither the tuning we observed, nor its relationship to behavioral performance. Instead, our results are well in line with previous work on higher-level spatial representations. In a naturalistic task such as ours, this necessarily includes sensory experiences. Our results support the notion that cognitive processes are closely intertwined with visual processing up to its earliest stages in the brain.

Influence of gaze direction

One limitation of the current study is that we cannot differentiate between vHD and the direction of gaze. Other studies using eye tracking indeed suggested that the landmarks capture most of the viewing (Navarro Schröder et al. 2018). Neurons that encode the allocentric location of gaze could

hence in theory lead to a directionality on population level if location and gaze are correlated. Such cells exist in the primate hippocampus and encode allocentric view independent from head direction (Rolls and Wirth, 2018; Rolls et al., 1997) or together with HD and self-location in a conjunctive code (Wirth et al., 2017). Computational models proposed that it is the conjunction between location and direction that mediates the environmental anchoring of HD in the RSC (Bicanski and Burgess, 2016). In monkeys (Killian et al., 2015) and humans (Julian et al., 2018; Nau et al., 2018c, 2018a; Staudigl et al., 2018), eye movement direction has also been shown to modulate activity in the entorhinal cortex in which we also observed a directional code. However, these studies focused on eye movements in the 2D plane (e.g. left-right, up-down) and the results are likely a manifestation of another grid-cell related mechanism (Killian et al., 2012; Nau et al., 2018a; Bicanski and Burgess, 2019). Since gaze direction typically varies more broadly over time than vHD, estimated tuning curves in visually responsive areas might broaden. This should have a stronger influence on visual compared to HD-representations, potentially explaining the relatively broad tuning in EVC (Fig. 5) as well as the influence of locomotion (Fig. 6). If correct, this predicted that early visual cortex should show a sharper tuning when analyzed as a function of gaze direction compared to vHD. However, neurons might also represent information in different reference frames depending on task demand (Chen et al., 2018). Future studies using eye tracking could address these questions directly. Acknowledging this ambiguity, we here refer to ‘directional tuning’ more generally.

Memory retrieval and learning

We hypothesized that participants with stronger directional tuning in higher-level visual and mnemonic regions should perform better in the spatial memory task. We based this prediction on the idea that stronger tuning indicated enhanced retrieval of directional information from memory. Our empirical test showed the opposite: participants with stronger tuning in higher-level mnemonic regions performed worse in the spatial memory task. Neither data quality or model parameters (Supplementary Fig. 3), nor directional sampling (Supplementary Fig. 1) could explain these results. While the underlying mechanism remains unclear, we can speculate about its nature. For that, we feel it is important to revisit what ‘tuning’ means in this context. The ‘tuning strength’ describes how well the time course of a voxel could be predicted. This does not necessarily constitute a net increase or decrease in activity, but instead how well we could predict the fluctuations over time. A voxel that weakly but consistently follows a directional modulation could hence show stronger tuning than one that strongly codes direction at very few time points.

If the tuning reflected memory encoding instead of retrieval, the tuning should be stronger in participants that kept improving in the task compared to participants who already reached ceiling.

Accordingly, we tested whether the high-memory-error group kept improving, and whether the low-memory-error group had already reached performance ceiling. We tested this post-hoc behaviorally, revealing that this was indeed the case. Both participant groups approached the same level of performance and participants with stronger tuning in higher-level regions showed steeper learning curves than those with weaker tuning (Supplementary Fig. 2C). A similar pattern of results has been reported for the hippocampus, whose activity tracks the amount of knowledge obtained at a given time rather than the accumulated absolute knowledge (Wolbers and Büchel, 2005).

In the present study, we did not investigate how directional representations develop over time. Yet, our observations are in line with previous work on the emergence of spatial representations in scene processing and navigation regions. For example, the RSC and PHG were shown to rapidly encode a novel environment by integrating information across different viewpoints and landmarks. Spatial representations emerging in the RSC were associated with a participant's ability to identify multiple scenes as belonging to the same or different location (Berens et al., 2019), as well as with wayfinding ability by registering landmark permanence (Auger et al., 2017). Consistent with the present results, the RSC activity could hence signal whether an environment has been successfully encoded or not. We observed a broader tuning in RSC in well compared to poorly performing participants, potentially indicating that the RSC processed information with a broader field of view. This could aid the integration of different landmarks and viewpoints (Auger et al., 2017; Berens et al., 2019). Importantly, the medial parietal lobe, which comprises the RSC, is known for both its perceptual and mnemonic capacities (O'Craven and Kanwisher, 2000), which might be topologically distributed. Previous work revealed an anterior-posterior distinction for scene construction and scene perception (Epstein and Baker, 2019; Silson et al., 2016, 2019). Our results (Fig. 4B) are consistent with this observation in two aspects. First, we also observed an anterior-posterior split on group-level in bilateral medial parietal lobe as a function of memory performance (Fig. 4B). Second, the participant group with more anterior tuning also performed better in the spatial memory task. Given above-mentioned reports, we speculate that this topology could indicate enhanced scene construction in participants with good memory performance. Again, this would also be consistent with the idea that these participants have already successfully encoded the environment and the object locations in it.

Notably, not only the RSC (Fig. 5C), but also the pmEC tuning (Fig. 5B) depended on the participants behavior, suggesting a different but related effect in pmEC. In line with this, rodent reorientation behavior depends on the stability of entorhinal HD-cell firing relative to the environment (Weiss et al., 2017). This alignment is also tightly coupled with the one of other spatial codes such as the hippocampal place field map (Knierim et al., 1995), which in turn predicts goal-oriented navigation behavior (Keinath et al., 2017). Since the entorhinal cortex is the key mediator

of hippocampal-cortical communication (Witter et al., 2017) and since the fMRI signal reflects input and local processing rather than output spiking (Logothetis et al., 2001), the present effects could therefore also reflect a related mechanism orchestrated by the hippocampus. This might include a reduction in hippocampal activity as environments become more familiar (Kaplan et al., 2014) or the successful encoding of object locations relative to the boundary (Doeller et al., 2008). Finally, the effects could also reflect individual preferences for egocentric versus allocentric navigation strategies (Gramann et al., 2010).

Future studies employing more dynamic model tests such as inverted encoding modeling (proof of principle in Supplementary Fig. 7) could address these questions directly by reconstructing the modeled behavioral features more dynamically (Gardner and Liu, 2019; Kriegeskorte and Douglas, 2019; Liu et al., 2018; Naselaris et al., 2011; Sprague et al., 2018, 2019). In addition, high temporal resolution measures such as magnetoencephalography (MEG) or intracranial recordings in combination with VR-tasks (Kunz et al., 2019) could monitor trial by trial changes in tuning as a function of behavioral performance.

Conclusion

Using virtual reality, a novel behavioral encoding model and 7T-fMRI, we examined world-centered directional tuning during active spatial behavior in humans. We demonstrated such tuning in visual, retrosplenial and parahippocampal cortices and the hippocampus. By mapping the tuning width across the cortex, we revealed a narrow-to-broad organization along the parahippocampal long-axis. Entorhinal and parahippocampal tuning strength as well as retrosplenial tuning width and topology reflected how well participants performed in a spatial memory task. We provide evidence that these effects likely depend on the encoding of the environment and the object locations within it. Finally, we show that the tuning in visual, retrosplenial and parahippocampal cortices as well as the hippocampus reflects the locomotory state. These results show the efficacy of encoding models for studying neural population dynamics during naturalistic navigation in human fMRI. They demonstrate a direct link between neural population coding and cognition and show that high-level cognitive processes modulate directional tuning in the service of behavior.

Author contributions

M.N., T.N.S. and C.F.D. conceived the experiment. T.N.S. acquired and preprocessed the data. M.N. analyzed the data and wrote the manuscript with input from M.F., T.N.S and C.F.D. All authors discussed the results and contributed to the manuscript.

Declaration of interest

The authors declare no conflicts of interest.

Acknowledgements

We thank Raphael Kaplan for helpful comments on an earlier version of this manuscript and Joshua B. Julian for helpful discussions. C.F.D.'s research is supported by the Max Planck Society, the Kavli Foundation, the European Research Council (ERC-CoG GEOCOG 724836), the Centre of Excellence scheme of the Research Council of Norway – Centre for Neural Computation (223262/F50), The Egil and Pauline Braathen and Fred Kavli Centre for Cortical Microcircuits, the National Infrastructure scheme of the Research Council of Norway – NORBRAIN (197467/F50), and the Netherlands Organisation for Scientific Research (NWO-Vidi 452-12-009, NWO-MaGW 406-14-114, and NWO-MaGW 406-15-291).

Methods

Participants

We recruited 26 participants for this study (11 females, 19-36 years old). Four subjects were excluded because of excessive head motion, i.e. the number of instantaneous movements larger than 0.5mm (Power et al., 2012) exceeded the across-participant average for more than one standard deviation. Another 2 participants were excluded because they finished fewer than four scanning runs. A total of 20 participants entered the analysis. The study was approved by the local research ethics committees (ethics committee University Duisburg-Essen, Germany and CMO region Arnhem-Nijmegen, NL) and participants gave written consent prior to scanning.

Virtual reality task

Participants performed a self-paced object-location memory task in virtual reality (Fig. 1A) adapted from Doeller and colleagues (Doeller et al., 2010). The circular virtual arena was created using the UnrealEngine2 Runtime software and was surrounded by 12 distinct landmarks positioned in steps of 30° and matched in visual similarity (triangles either tilted up- or downwards, with red, green and blue colored corners). Participants could freely navigate in this arena via key-presses. The smallest instantaneous rotational movement possible was 10° and translational movement speed was constant after a 500ms ramp. In the beginning of the experiment, six everyday objects had to be collected, which were scattered across the arena. Across different trials and without the objects being present, participants were prompted to navigate to the location of a previously cued object. After indicating the remembered location via key press (drop), the respective object appeared at the correct location to give feedback and participant collected the object again before the next trial began. After an average of 3 trials (range 2-4), a fixation cross was presented on a gray background for 4 seconds. An average of 179 trials were performed (range: 94-253 trials due to the self-paced nature of the task, Supplementary Fig. 2D) and object locations were randomized across participants. In order to explain the task and to familiarize participants with it, they performed a similar task with different objects in a different virtual environment prior to scanning. We tracked the improvement in memory performance over trials by assessing the ‘memory error’, i.e. the Euclidean distance between true and remembered location in each trial measured in virtual vertices (arbitrary units).

Behavioral analysis

To ensure that there were no prominent or distinct directional cues that biased navigation behavior, we ruled out differences in the time spent facing into different directions. Supplementary Fig. 1 depicts the results of these analyses for all participants and time points, split into locomotion and stationary periods as well as into high- and low-memory-error participant groups. We accounted for individual differences in absolute time spent in the experiment by expressing time spent as percent

of the total experimental duration. We binned vHD in steps of 10° and performed a repeated-measure (rm) ANOVA across directions, which did not reveal any biases in directional sampling ($F(35, 665) = 0.77$, $p = 0.834$), also not when testing the high-memory error group ($F(35, 315) = 0.55$, $p = 0.984$) or the low-memory error group ($F(35, 315) = 0.87$, $p = 0.675$) individually. In addition, we did not observe biases in directional sampling across the experiment when splitting the data into locomotion ($F(35, 665) = 0.79$, $p = 0.806$) and stationary ($F(35, 665) = 0.87$, $p = 0.681$) periods (Supplementary Fig. 1A).

In addition to the directional sampling in the course of the experiment, we analyzed the distribution of vHD within each TR. We again binned vHD into 10° steps and converted it into percent of total viewing time. We then circular-shifted each of the resulting histograms such that the most sampled direction lined up across TRs (Φ), revealing that participants spent 52% of the time within each TR facing into a single direction (Fig. 1B, Supplementary Fig. 1A). The distribution of vHD within each TR was therefore non-uniform and centered on one predominant direction. We used a two-tailed permutation-based unpaired t-test to compare the time spent facing towards this predominant direction within each TR across across groups, which did not reveal a difference ($t(18) = -1.26$, $p = 0.224$, $k = 10000$, Supplementary Fig. 1C). Using a paired version of this test, we did observed a difference between stationary and locomotion periods ($t(19) = 6.12$, $p = 0.0001$, $k = 10000$). Importantly, this had a differential effect for different regions of interests (Fig. 6), suggesting that there was no general positive or negative effect on model performance. During this task, participants spent around 54% of their time navigating, with shorter time spent rotating during locomotion compared to stationary periods (two-tailed permutation-based paired ttest: $t(38) = 7.95$, $p = 0.0001$, $k = 10000$). There were no differences in the time spent translating ($t(18) = 0.41$, $p = 0.690$, $k = 10000$) or rotating between participant groups (during locomotion: $t(18) = 0.55$, $p = 0.625$, $k = 10000$ and during stationary periods: $t(18) = 1.00$, $p = 0.338$, $k = 10000$).

MRI acquisition

During the object-location memory task in VR we acquired T2*-weighted functional images on a 7T Siemens MAGNETOM scanner using a 3D-EPI pulse sequence (Poser, Koopmans, Witzel, Wald, & Barth, 2010), a 32-channel head coil and following parameters: TR = 2756 ms, TE = 20 ms, flip angle = 14°, voxel size = 0.9 mm x 0.9 mm, slice thickness = 0.92 mm, slice oversampling = 8.3 %, 96 slices with a 210 mm x 210 mm field of view, phase encoding acceleration factor = 4, 3D acceleration factor = 2. The first 5 volumes of each run were discarded. Functional images were acquired across 5 scanning runs of 210 TR's or approximately 10 minutes each. In addition, we acquired T1-weighted structural images (MP2RAGE; voxel size: 0.63 mm isotropic) and a B0-field map (gradient echo; voxel size: 1.8 x 1.8 x 2.2 mm3) for each participant.

Preprocessing

The data used here were used in two previous reports (Navarro Schröder et al., 2015, 2018). Data were preprocessed using the automatic analysis library (Cusack et al., 2015), utilizing functions of several analysis packages. For each participant, functional images were realigned using SPM8, followed by independent component analysis (ICA) denoising using FIX artifact removal implemented in FSL 5.0.4. To improve signal-to-noise ratio, and with it the ICA-detection of noise components, data were smoothed with a Gaussian full-width-at-half-maximum kernel of 2.5 mm. Images were then non-linearly normalized to a group-average EPI-template using the Advanced Neuroimaging Toolbox (ANTs) (Avants et al., 2011) and high-pass filtered with a 128-s cutoff using FSL. Voxel-wise variance explained by the six realignment parameters (x,y,z, pitch, roll, yaw) as well as by spikes (sudden deviations in signal intensity of more than two temporal standard deviations) was removed via nuisance regression. Out-of-brain voxels were excluded.

Regions of interests (ROIs)

In a ROI analysis we tested human scene processing and navigation regions (Fig. 5A) that were previously proposed to support cognitive mapping (Byrne et al., 2007). The hippocampal (HPC), anterolateral entorhinal (alEC) and posteromedial entorhinal (pmEC) ROIs were defined manually using ItK-SNAP (www.itksnap.org) based on the high-resolution group average EPI-template. The entorhinal masks were based on previous reports (Navarro Schröder et al., 2015), in which the entorhinal mask was divided into anterolateral (alEC) and posteromedial entorhinal cortex (pmEC). The ROIs for the parahippocampal gyrus (PHG) as well as the retrosplenial cortex (RSC) were based on the reverse inference meta - analysis for “Retrosplenial cortex” and “Parahippocampal cortex” using Neurosynth (Yarkoni et al., 2011). We took the top 5% highest probability voxels from each respective Neurosynth map and removed isolated voxels from the resulting binary masks. This procedure resulted in coherent bilateral clusters in the medial parietal cortex and parahippocampal gyrus respectively. The early visual cortex (EVC) ROI was created by thresholding the corresponding probability map ‘Visual_hOc1’ of the SPM anatomy toolbox at 50% and co-registering it non-linearly to our group-average template space. To do so, we used SPM to first segment the group template and then to ‘normalize’ the ROI into our template-space using the resulting tissue maps and nearest-neighbor interpolation. The resulting ROI masks were located at following average MNI coordinates [X,Y,Z] and were of following size. EVC: left hemisphere [-4, -88, 0], right hemisphere [14, -86, 0], n voxels: 16299, RSC: [-14, -56, 12] and [18, -54, 14], n voxels: 1926, PHG: [-26, -38, -12] and [26, -34, -16], n voxels: 2392, HPC: [-24, -24, -14] and [28, -22, -14], n voxels: 9781, alEC: [-20, 0, -34] and [22, 0, -34], n voxels: 1693, pmEC: [-20, -10, -28] and [20, -8, -28], n voxels: 1692.

Analysis overview

Our analysis estimated the directional tuning of a voxel in several steps. First, we built a vHD-encoding model by incorporating the participant's navigation behavior into basis sets of circular-gaussian von-Mises distributions, which we call vHD-kernels. Each individual direction was modeled with a different vHD-kernel, each representing a smooth directional tuning without discretizing the data into bins. Second, we estimated voxel-wise weights for each of these kernels, together representing a voxel's tuning curve. We refer to this step as model training. Third, we used these weights to predict activity in held out data which constituted the model test. This way we obtained a measure of model performance for the given vHD basis set. Finally, by iteratively varying the full-width-at-half-maximum of the vHD-kernels in the basis set and repeating above mentioned steps, we not only tested one vHD-basis set, but multiple ones. This approach allowed us to also estimate the tuning width of each voxel (the kernel width that maximized prediction accuracy). All processing steps mentioned comprised several individual sub-steps, which are described in detail below.

Building the virtual head direction (vHD) encoding model

We modeled vHD using a basis set of circular-Gaussian von Mises distributions as implemented in the Circular Statistics Toolbox for Matlab (Berens, 2009). Each kernel in this basis set covered the full 360° with an angular resolution of 1°. Across different iterations of our analysis, we varied the full-width-at-half-maximum of these kernels, each time testing how well the resulting model weights allowed to predict activity in held-out data. To balance directional sensitivity across iterations, the spacing between kernels always matched the kernel width (i.e. the broader each individual kernel, the fewer kernels were used). We tested following kernel widths, all representing divisibles of 360°: 10°, 15°, 20°, 24°, 30°, 36°, 45°, 60°. Note that our regions of interest analysis builds on weight-shuffling to rule out that the number of kernels influenced model accuracies. For each kernel and participant, we computed the predicted kernel activities based on the vHD over time. To give an example, for a given vHD of 30° a kernel centered on 30° was assigned high activity, a kernel centered on 40° slightly lower activity and a kernel centered on 200° was assigned very low activity. By doing this for all time points and kernels, we built regressors representing the predicted activity given a directional tuning and the specific vHD over time. Since vHD was sampled at higher temporal resolution than the imaging data, we then computed the within-TR activity of each kernel as the median activity across all time points within the TR. Finally, the resulting regressors were scaled from 0 to 1 and convolved with the hemodynamic response function (HRF) as implemented in SPM12 (https://www.fil.ion.ucl.ac.uk/spm/software/spm12/) using default settings (kernel length: 32 s, time to peak: 6s). Each regressor represented a predicted activity profile over time as modeled by the respective kernel. To follow the example above, the activity of a voxel encoding the direction 30° should be more similar to the activity predicted by the kernel centered on 30°, than to the one predicted by the kernel centered on 200°.

Model training

We estimated voxel-wise weights for all vHD-kernels in the corresponding basis set using l2-regularized (ridge) regression. To improve the directionality of this model, we added a covariate modeling movement independent of direction whose weight was discarded. Because vHD is not independent at two successive time points, the resulting design was multicollinear. Ridge regression avoids potential biases in the resulting weights by penalizing high coefficients, which could otherwise affect model accuracies. Since the regularization parameter (λ) cannot be known a priori, our model training builds on leave-one-out cross-validation to find the optimal λ and with it the optimal model weights. As training set, we used the first two and the last two scanning runs, leaving the third run as the final and independent test set. A scan run typically took around 10 minutes. Since run 3 was acquired in the middle of the experiment, our results are invariant to the duration of the scanning session. If a participant did not complete all five runs, we always used the third valid run as a test set and all others as training set. Within the training set, we used all runs except one to fit voxel-wise weights for ten different regularization parameters log-spaced between 1 to 10,000,000, each time testing how well these weights predicted the activity in the left-out validation run. To assess prediction performance, we used Pearson correlation between the real time course of a voxel in the validation set and the time course predicted by vHD weighted by the estimated model weights. We cross-validated this prediction such that each run within the training set served as validation set once. The regularization parameter that led to the best prediction performance on average was then determined for each voxel. If no clear best-performing λ could be determined (i.e. Pearson's R negatively approximated zero with increasing regularization), the respective voxel was excluded. We then averaged λ across voxels within each participant and used it to estimate the final model weights using the full training set. These model weights serve as basis for all further model tests described below.

Model test

All model tests were performed on the held-out and independent test set (one scan run of around 10 minutes in the middle of the experiment). First, we predicted voxel-wise activity (similar to the model training) in a univariate 'forward-model' approach. For each voxel, we generated a predicted time course by weighting the design matrix of the test run after HRF-convolution by the model weights obtained for this voxel during model training. The resulting predicted time course was then compared to the observed time course using Pearson correlation. Please note that all steps described below were repeated for multiple vHD-basis sets differing in the number and width of the corresponding vHD kernels.

We mapped directional tuning across the cortex using the statistical non-parametric mapping (SnPM) toolbox ((Nichols and Holmes, 2002), http://warwick.ac.uk/snpm). We performed a permutation-based one-sample t-test of model performance (Pearson correlations) against zero (k =

10000 shuffles, input image and variance smoothing: 7.2mm). To reduce computational costs the preprocessed data were downsampled from 0.9 mm to 1.8 mm isotropic for this step. This was repeated for each vHD basis set. We then used the SnPM toolbox to threshold the resulting pseudo-T-maps at a FDR-corrected $p<0.05$. For each voxel we then selected the across-participant median tuning width of the vHD basis set that maximized the pseudo-T. Figure 3 depicts these results for all participants. For visualization we plot the results overlaid on the group-average T1 template at T1-resolution obtained via nearest-neighbor interpolation. We repeated this analysis split into high-memory-error and low-memory error participants (i.e. median-split of memory error, Fig. 4). Note that each participant group comprised $n = 10$ participants, resulting in 1024 possible shuffles and a minimal possible $p = 0.000977$. Because this precludes FDR-correction we use uncorrected $p = 0.001$ to visualize these sub-group effects.

In addition to the voxel-wise group analysis we also conducted a regions of interest analysis for areas involved in scene processing and navigation such as the early visual cortex, retrosplenial cortex, parahippocampal cortex, the posteromedial entorhinal cortex and the hippocampus (Fig. 5A). We again performed model training and test, this time focusing specifically on voxels in our ROIs. This greatly reduced the number of voxels and hence computational cost. To avoid potential influences of the number of kernels on these ROI results, we instead performed voxel-wise bootstrapping to convert the Pearson correlations into z-scores. The necessary null distribution of each voxel was obtained by shuffling the training weights 500 times, each time computing Pearson's R between the predicted and the observed time course of the test set. All shuffles were unique. Since not all voxels in our (probabilistic) ROIs were expected to carry vHD-information and to increase robustness of our effects, we performed voxel selection within each ROI, limiting the model test to voxels with high predictability in the model training (top 25 % highest prediction accuracy in the training). If a voxel was not directionally tuned in the training, we did not expect it to be directionally tuned in the test. Finally, the z-scores of the remaining voxels were averaged within each ROI. Again, we iterated this analysis for all basis sets (Fig. 5B), yielding the tuning strength (model performance) and the tuning width (the width of the vHD-kernels in the basis set) of each ROI. Statistical inference was performed for each ROI and participant group (high- and low-memory-error participants) using permutation-based one-sample t-tests on group level (all possible 1024 shuffles, $n = 2 \times 10$) as implemented in the mult_comp_perm_t1 function distributed by Mathworks (https://se.mathworks.com/matlabcentral/fileexchange).
To test for differences across participant groups, we used permutation-based unpaired two-sample two-tailed t-tests ($k = 10000$) as implemented in statcond distributed via the EEGLab Matlab toolbox. (https://github.com/openroc/eeglab/blob/master/tags/EEGLAB7_0_0_0beta).

In addition to the forward model test, we also inverted the encoding model to reconstruct, or 'decode', vHD multivarietly from the population of voxels within each ROI (Supplementary Fig. 7). We multiplied the Moore-Penrose pseudoinverse of all voxel-wise weights in an ROI (m voxels x k weights)$^{-1}$ with the multivoxel-pattern at each image acquisition (m voxels) to obtain the estimated vHD-kernel activities at each for volume (k weights). By doing this for all volumes, we reconstructed the vHD-kernel activities for the entire test set of each participant. To assess reconstruction performance, we used 2D correlation between the reconstructed kernel activities (k weights x n TR's) and the design matrix of the test run (also k weights x n TR's). Since again the number of kernels and hence reconstructed weights differed across iterations and basis sets, we used weight-shuffling to convert the resulting correlations into z-scores. As in the forward model, we shuffled the model weights (500 random unique shuffles), each time going through the full vHD-reconstruction procedure. As expected, the results obtained by inverting the encoding model (Supplementary Fig. 7) resemble the ones obtained by the voxel-wise forward-model procedure (Fig. 5).

<u>Temporal signal-to-noise ratio (tSNR) and model regularization do not explain the results</u>

The estimated optimal regularization parameter λ (Supplementary Fig. 3A) depended on the basis set (Supplementary Fig. 3B) (rmANOVA results: $F(7, 126) = 11.80$, $p = 1.7 \times 10\text{-}11$) as expected, but not on participant group ($F(1, 18) = 0.0037$, $p = 0.952$) and there was no interaction between the two ($F(7,126) = 0.37$, $p = 0.920$). There were differences in tSNR across ROIs (rmANOVA results: $F(5, 90) = 258.29$, $p = 9.1 \times 10\text{-}52$), but not across participant groups ($F(1, 18) = 0.082$, $p = 0.777$) and there was no interaction between the two ($F(5, 90) = 1.34$, $p = 0.255$) (Supplementary Fig. 3C). Neither model performance (Spearman correlation: rho = 0.076, p = 0.409), nor tuning width (rho = -0.059, p = 0.525) correlated with tSNR Supplementary Fig. 3D.

References

Acharya, L., Aghajan, Z.M., Vuong, C., Moore, J.J., and Mehta, M.R. (2016). Causal Influence of Visual Cues on Hippocampal Directional Selectivity. Cell *164*, 197–207.

Aguirre, G.K. (1999). Topographical disorientation: a synthesis and taxonomy. Brain *122*, 1613–1628.

Auger, S.D., Zeidman, P., and Maguire, E.A. (2017). Efficacy of navigation may be influenced by retrosplenial cortex-mediated learning of landmark stability. Neuropsychologia *104*, 102–112.

Avants, B.B., Tustison, N.J., Song, G., Cook, P.A., Klein, A., and Gee, J.C. (2011). A reproducible evaluation of ANTs similarity metric performance in brain image registration. NeuroImage *54*, 2033–2044.

Ayaz, A., Saleem, A.B., Schölvinck, M.L., and Carandini, M. (2013). Locomotion Controls Spatial Integration in Mouse Visual Cortex. Curr. Biol. *23*, 890–894.

Baumann, O., and Mattingley, J.B. (2010). Medial Parietal Cortex Encodes Perceived Heading Direction in Humans. J. Neurosci. *30*, 12897–12901.

Bellmund, J.L., Deuker, L., Navarro Schröder, T., and Doeller, C.F. (2016). Grid-cell representations in mental simulation. ELife *5*.

Berens, P. (2009). CircStat: A MATLAB Toolbox for Circular Statistics. J. Stat. Softw. *31*.

Berens, S.C., Joensen, B.H., and Horner, A.J. (2019). Tracking the emergence of location-based spatial representations. BioRxiv.

Bicanski, A., and Burgess, N. (2016). Environmental Anchoring of Head Direction in a Computational Model of Retrosplenial Cortex. J. Neurosci. *36*, 11601–11618.

Bicanski, A., and Burgess, N. (2018). A neural-level model of spatial memory and imagery. ELife *7*.

Bicanski, A., and Burgess, N. (2019). A Computational Model of Visual Recognition Memory via Grid Cells. Curr. Biol. *29*, 979-990.e4.

Bosch, S.E., Jehee, J.F.M., Fernandez, G., and Doeller, C.F. (2014). Reinstatement of Associative Memories in Early Visual Cortex Is Signaled by the Hippocampus. J. Neurosci. *34*, 7493–7500.

Butler, W.N., Smith, K.S., van der Meer, M.A.A., and Taube, J.S. (2017). The Head-Direction Signal Plays a Functional Role as a Neural Compass during Navigation. Curr. Biol. *27*, 1259–1267.

Byrne, P., Becker, S., and Burgess, N. (2007). Remembering the past and imagining the future: a neural model of spatial memory and imagery. Psychol. Rev. *114*, 340–375.

Chadwick, M.J., Jolly, A.E.J., Amos, D.P., Hassabis, D., and Spiers, H.J. (2015). A Goal Direction Signal in the Human Entorhinal/Subicular Region. Curr. Biol. *25*, 87–92.

Chen, L.L., Lin, L.-H., Green, E.J., Barnes, C.A., and McNaughton, B.L. (1994). Head-direction cells in the rat posterior cortex: I. anatomical distribution and behavioral modulation. Exp. Brain Res. *101*, 8–23.

Chen, X., DeAngelis, G.C., and Angelaki, D.E. (2018). Flexible egocentric and allocentric representations of heading signals in parietal cortex. Proc. Natl. Acad. Sci. *115*, E3305–E3312.

Cheng, K. (1986). A purely geometric module in the rat's spatial representation. Cognition *23*, 149–178.

Cho, J., and Sharp, P.E. (2001). Head direction, place, and movement correlates for cells in the rat retrosplenial cortex. Behav. Neurosci. *115*, 3–25.

Clark, B.J., Simmons, C.M., Berkowitz, L.E., and Wilber, A.A. (2018). The Retrosplenial-Parietal Network and Reference Frame Coordination for Spatial Navigation. PsyArXiv Prepr. 1–35.

Coletta, S., Frey, M., Nasr, K., Preston-Ferrer, P., and Burgalossi, A. (2018). Testing the Efficacy of Single-Cell Stimulation in Biasing Presubicular Head Direction Activity. J. Neurosci. *38*, 3287–3302.

Cullen, K.E., and Taube, J.S. (2017). Our sense of direction: progress, controversies and challenges. Nat. Neurosci. *20*, 1465–1473.

Cusack, R., Vicente-Grabovetsky, A., Mitchell, D.J., Wild, C.J., Auer, T., Linke, A.C., and Peelle, J.E. (2015). Automatic analysis (aa): efficient neuroimaging workflows and parallel processing using Matlab and XML. Front. Neuroinformatics *8*.

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

Doeller, C.F., Barry, C., and Burgess, N. (2010). Evidence for grid cells in a human memory network. Nature *463*, 657–661.

Dudchenko, P.A., Wood, E.R., and Smith, A. (2019). A new perspective on the head direction cell system and spatial behavior. Neurosci. Biobehav. Rev.

Dumoulin, S.O., and Wandell, B.A. (2007). Population receptive field estimates in human visual cortex. NeuroImage *39*, 647–660.

Epstein, R., and Kanwisher, N. (1998). A cortical representation of the local visual environment. Nature *392*, 598–601.

Epstein, R.A., and Baker, C.I. (2019). Scene Perception in the Human Brain. Annu. Rev. Vis. Sci. *5*.

Epstein, R.A., Patai, E.Z., Julian, J.B., and Spiers, H.J. (2017). The cognitive map in humans: spatial navigation and beyond. Nat. Neurosci. *20*, 1504–1504.

Ferrara, K., and Park, S. (2016). Neural representation of scene boundaries. Neuropsychologia *89*, 180–190.

Fournier, J., Saleem, A.B., Diamanti, E.M., Wells, M.J., Harris, K.D., and Carandini, M. (2019). Modulation of visual cortex by hippocampal signals. BioRxiv.

Galletti, C., Squatrito, S., Battaglini, P.P., and Grazia Maioli, M. (1984). "Real-motion" cells in the primary visual cortex of macaque monkeys. Brain Res. *301*, 95–110.

Gardner, J.L., and Liu, T. (2019). Inverted Encoding Models Reconstruct an Arbitrary Model Response, Not the Stimulus. Eneuro *6*, ENEURO.0363-18.2019.

Giocomo, L.M., Stensola, T., Bonnevie, T., Van Cauter, T., Moser, M.-B., and Moser, E.I. (2014). Topography of Head Direction Cells in Medial Entorhinal Cortex. Curr. Biol. *24*, 252–262.

Gourtzelidis, P., Tzagarakis, C., Lewis, S.M., Crowe, D.A., Auerbach, E., Jerde, T.A., Uğurbil, K., and Georgopoulos, A.P. (2005). Mental maze solving: directional fMRI tuning and population coding in the superior parietal lobule. Exp. Brain Res. *165*, 273–282.

Gramann, K., Onton, J., Riccobon, D., Mueller, H.J., Bardins, S., and Makeig, S. (2010). Human Brain Dynamics Accompanying Use of Egocentric and Allocentric Reference Frames during Navigation. J. Cogn. Neurosci. *22*, 2836–2849.

Hartley, T., Lever, C., Burgess, N., and O'Keefe, J. (2014). Space in the brain: how the hippocampal formation supports spatial cognition. Philos. Trans. R. Soc. B Biol. Sci. *369*, 20120510.

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

Huth, A.G., de Heer, W.A., Griffiths, T.L., Theunissen, F.E., and Gallant, J.L. (2016). Natural speech reveals the semantic maps that tile human cerebral cortex. Nature *532*, 453–458.

Indovina, I., Maffei, V., Pauwels, K., Macaluso, E., Orban, G.A., and Lacquaniti, F. (2013). Simulated self-motion in a visual gravity field: Sensitivity to vertical and horizontal heading in the human brain. NeuroImage *71*, 114–124.

Jacob, P.-Y., Casali, G., Spieser, L., Page, H., Overington, D., and Jeffery, K. (2017). An independent, landmark-dominated head-direction signal in dysgranular retrosplenial cortex. Nat. Neurosci. *20*, 173–175.

Jacobs, J., Kahana, M.J., Ekstrom, A.D., Mollison, M.V., and Fried, I. (2010). A sense of direction in human entorhinal cortex. Proc. Natl. Acad. Sci. *107*, 6487–6492.

Julian, J.B., Keinath, A.T., Muzzio, I.A., and Epstein, R.A. (2015). Place recognition and heading retrieval are mediated by dissociable cognitive systems in mice. Proc. Natl. Acad. Sci. *112*, 6503–6508.

Julian, J.B., Keinath, A.T., Frazzetta, G., and Epstein, R.A. (2018). Human entorhinal cortex represents visual space using a boundary-anchored grid. Nat. Neurosci.

Kaplan, R., Horner, A.J., Bandettini, P.A., Doeller, C.F., and Burgess, N. (2014). Human hippocampal processing of environmental novelty during spatial navigation: Human Hippocampal Processing Of Environmental Novelty. Hippocampus *24*, 740–750.

Kay, K.N., Naselaris, T., Prenger, R.J., and Gallant, J.L. (2008). Identifying natural images from human brain activity. Nature *452*, 352–355.

Keinath, A.T., Julian, J.B., Epstein, R.A., and Muzzio, I.A. (2017). Environmental Geometry Aligns the Hippocampal Map during Spatial Reorientation. Curr. Biol. *27*, 309–317.

Killian, N.J., Jutras, M.J., and Buffalo, E. a. (2012). A map of visual space in the primate entorhinal cortex. Nature *491*, 761–764.

Killian, N.J., Potter, S.M., and Buffalo, E.A. (2015). Saccade direction encoding in the primate entorhinal cortex during visual exploration. Proc. Natl. Acad. Sci. U. S. A. *112*, 15743–15748.

Kim, M., and Maguire, E.A. (2019). Encoding of 3D head direction information in the human brain. Hippocampus *29*, 619–629.

Knierim, J., Kudrimoti, H., and McNaughton, B. (1995). Place cells, head direction cells, and the learning of landmark stability. J. Neurosci. *15*, 1648–1659.

Kok, P., and Turk-Browne, N.B. (2018). Associative Prediction of Visual Shape in the Hippocampus. J. Neurosci. *38*, 6888–6899.

Kornienko, O., Latuske, P., Bassler, M., Kohler, L., and Allen, K. (2018). Non-rhythmic head-direction cells in the parahippocampal region are not constrained by attractor network dynamics. ELife *7*.

Kravitz, D.J., Peng, C.S., and Baker, C.I. (2011). Real-World Scene Representations in High-Level Visual Cortex: It's the Spaces More Than the Places. J. Neurosci. *31*, 7322–7333.

Kriegeskorte, N., and Douglas, P.K. (2019). Interpreting encoding and decoding models. Curr. Opin. Neurobiol. *55*, 167–179.

Kunz, L., Maidenbaum, S., Chen, D., Wang, L., Jacobs, J., and Axmacher, N. (2019). Mesoscopic Neural Representations in Spatial Navigation. Trends Cogn. Sci. *23*, 615–630.

Lescroart, M.D., and Gallant, J.L. (2019). Human Scene-Selective Areas Represent 3D Configurations of Surfaces. Neuron *101*, 178-192.e7.

Liu, T., Cable, D., and Gardner, J.L. (2018). Inverted Encoding Models of Human Population Response Conflate Noise and Neural Tuning Width. J. Neurosci. *38*, 398–408.

Logothetis, N.K., Pauls, J., Augath, M., Trinath, T., and Oeltermann, a (2001). Neurophysiological investigation of the basis of the fMRI signal. Nature *412*, 150–157.

Maass, A., Berron, D., Libby, L.A., Ranganath, C., and Düzel, E. (2015). Functional subregions of the human entorhinal cortex. ELife *4*, 1–20.

Marchette, S.A., Vass, L.K., Ryan, J., and Epstein, R.A. (2014). Anchoring the neural compass: Coding of local spatial reference frames in human medial parietal lobe. Nat. Neurosci. *17*, 1598–1606.

van der Meer, M.A.A., Richmond, Z., Braga, R.M., Wood, E.R., and Dudchenko, P.A. (2010). Evidence for the use of an internal sense of direction in homing. Behav. Neurosci. *124*, 164–169.

Meister, M.L.R., and Buffalo, E.A. (2016). Getting directions from the hippocampus: The neural connection between looking and memory. Neurobiol. Learn. Mem. *134*, 135–144.

Mitchell, A.S., Czajkowski, R., Zhang, N., Jeffery, K., and Nelson, A.J.D. (2018). Retrosplenial cortex and its role in spatial cognition. Brain Neurosci. Adv. *2*, 239821281875709.

Mitchell, T.M., Shinkareva, S.V., Carlson, A., Chang, K.-M., Malave, V.L., Mason, R.A., and Just, M.A. (2008). Predicting Human Brain Activity Associated with the Meanings of Nouns. Science *320*, 1191–1195.

Moser, E.I., Moser, M.-B., and McNaughton, B.L. (2017). Spatial representation in the hippocampal formation: a history. Nat. Neurosci. *20*, 1448–1464.

Naselaris, T., Kay, K.N., Nishimoto, S., and Gallant, J.L. (2011). Encoding and decoding in fMRI. NeuroImage *56*, 400–410.

Nau, M., Julian, J.B., and Doeller, C.F. (2018a). How the Brain's Navigation System Shapes Our Visual Experience. Trends Cogn. Sci. *22*, 810–825.

Nau, M., Schindler, A., and Bartels, A. (2018b). Real-motion signals in human early visual cortex. NeuroImage *175*, 379–387.

Nau, M., Navarro Schröder, T., Bellmund, J.L.S., and Doeller, C.F. (2018c). Hexadirectional coding of visual space in human entorhinal cortex. Nat. Neurosci.

Navarro Schröder, T., Haak, K.V., Zaragoza Jimenez, N.I., Beckmann, C.F., and Doeller, C.F. (2015). Functional topography of the human entorhinal cortex. ELife *4*, 1–17.

Navarro Schröder, T., Towse, B.W., Nau, M., Burgess, N., Barry, C., and Doeller, C.F. (2018). Entorhinal cortex minimises uncertainty for optimal behaviour. BioRxiv.

Nichols, T.E., and Holmes, A.P. (2002). Nonparametric permutation tests for functional neuroimaging: A primer with examples. Hum. Brain Mapp. *15*, 1–25.

O'Craven, K.M., and Kanwisher, N. (2000). Mental Imagery of Faces and Places Activates Corresponding Stimulus-Specific Brain Regions. J. Cogn. Neurosci. *12*, 1013–1023.

O'Keefe, J., and Nadel, L. (1978). The hippocampus as a cognitive map (Oxford : New York: Clarendon Press ; Oxford University Press).

Power, J.D., Barnes, K.A., Snyder, A.Z., Schlaggar, B.L., and Petersen, S.E. (2012). Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. NeuroImage *59*, 2142–2154.

Rademaker, R.L., Chunharas, C., and Serences, J.T. (2019). Coexisting representations of sensory and mnemonic information in human visual cortex. Nat. Neurosci.

Robertson, R.G., Rolls, E.T., Georges-Francois, P., and Panzeri, S. (1999). Head direction cells in the primate pre-subiculum. Hippocampus *9*, 206–219.

Rolls, E.T., and Wirth, S. (2018). Spatial representations in the primate hippocampus, and their functions in memory and navigation. Prog. Neurobiol. *171*, 90–113.

Rolls, E.T., Robertson, R.G., and Georges-François, P. (1997). Spatial View Cells in the Primate Hippocampus. Eur. J. Neurosci. *9*, 1789–1794.

Saleem, A.B., Ayaz, A., Jeffery, K.J., Harris, K.D., and Carandini, M. (2013). Integration of visual motion and locomotion in mouse visual cortex. Nat. Neurosci. *16*, 1864–1869.

Saleem, A.B., Diamanti, E.M., Fournier, J., Harris, K.D., and Carandini, M. (2018). Coherent encoding of subjective spatial position in visual cortex and hippocampus. Nature *562*, 124–127.

Sargolini, F., Fyhn, M., Hafting, T., McNaughton, B.L., Witter, M.P., Moser, M.-B., and Moser, E.I. (2006). Conjunctive representation of position, direction, and velocity in entorhinal cortex. Science *312*, 758–762.

Schindler, A., and Bartels, A. (2013). Parietal Cortex Codes for Egocentric Space beyond the Field of View. Curr. Biol. *23*, 177–182.

Shine, J.P., Valdés-Herrera, J.P., Hegarty, M., and Wolbers, T. (2016). The Human Retrosplenial Cortex and Thalamus Code Head Direction in a Global Reference Frame. J. Neurosci. *36*, 6371–6381.

Shine, J.P., Valdés-Herrera, J.P., Tempelmann, C., and Wolbers, T. (2019). Evidence for allocentric boundary and goal direction information in the human entorhinal cortex and subiculum. Nat. Commun. *10*.

Silson, E.H., Steel, A.D., and Baker, C.I. (2016). Scene-Selectivity and Retinotopy in Medial Parietal Cortex. Front. Hum. Neurosci. *10*.

Silson, E.H., Gilmore, A.W., Kalinowski, S.E., Steel, A., Kidder, A., Martin, A., and Baker, C.I. (2019). A Posterior–Anterior Distinction between Scene Perception and Scene Construction in Human Medial Parietal Cortex. J. Neurosci. *39*, 705–717.

Skaggs, W.E., Knierim, J.J., Kudrimoti, H.S., and McNaughton, B.L. (1995). A model of the neural basis of the rat's sense of direction. Adv. Neural Inf. Process. Syst. *7*, 173–180.

Sprague, T.C., Ester, E.F., and Serences, J.T. (2016). Restoring Latent Visual Working Memory Representations in Human Cortex. Neuron *91*, 694–707.

Sprague, T.C., Adam, K.C.S., Foster, J.J., Rahmati, M., Sutterer, D.W., and Vo, V.A. (2018). Inverted Encoding Models Assay Population-Level Stimulus Representations, Not Single-Unit Neural Tuning. Eneuro *5*, ENEURO.0098-18.2018.

Sprague, T.C., Boynton, G.M., and Serences, J.T. (2019). Inverted encoding models estimate sensible channel responses for sensible models. BioRxiv.

Staudigl, T., Leszczynski, M., Jacobs, J., Sheth, S.A., Schroeder, C.E., Jensen, O., and Doeller, C.F. (2018). Hexadirectional Modulation of High-Frequency Electrophysiological Activity in the Human Anterior Medial Temporal Lobe Maps Visual Space. Curr. Biol. *28*, 3325-3329.e4.

Taube, J.S. (2007). The Head Direction Signal: Origins and Sensory-Motor Integration. Annu. Rev. Neurosci. *30*, 181–207.

Taube, J.S., and Burton, H.L. (1995). Head direction cell activity monitored in a novel environment and during a cue conflict situation. J. Neurophysiol. *74*, 1953–1971.

Taube, J., Muller, R., and Ranck, J. (1990). Head-direction cells recorded from the postsubiculum in freely moving rats. I. Description and quantitative analysis. J. Neurosci. *10*, 420–435.

Thirion, B., Duchesnay, E., Hubbard, E., Dubois, J., Poline, J.-B., Lebihan, D., and Dehaene, S. (2006). Inverse retinotopy: Inferring the visual content of images from brain activation patterns. NeuroImage *33*, 1104–1116.

Valerio, S., and Taube, J.S. (2012). Path integration: how the head direction signal maintains and corrects spatial orientation. Nat. Neurosci. *15*, 1445–1453.

Vann, S.D., Aggleton, J.P., and Maguire, E.A. (2009). What does the retrosplenial cortex do? Nat. Rev. Neurosci. *10*, 792–802.

Vass, L.K., and Epstein, R.A. (2013). Abstract Representations of Location and Facing Direction in the Human Brain. J. Neurosci. *33*, 6133–6142.

Vass, L.K., and Epstein, R.A. (2016). Common Neural Representations for Visually Guided Reorientation and Spatial Imagery. Cereb. Cortex bhv343.

Wandell, B.A., and Winawer, J. (2015). Computational neuroimaging and population receptive fields. Trends Cogn. Sci. *19*, 349–357.

Weiss, S., and Derdikman, D. (2018). Role of the head-direction signal in spatial tasks: when and how does it guide behavior? J. Neurophysiol. *120*, 78–87.

Weiss, S., Talhami, G., Gofman-Regev, X., Rapoport, S., Eilam, D., and Derdikman, D. (2017). Consistency of Spatial Representations in Rat Entorhinal Cortex Predicts Performance in a Reorientation Task. Curr. Biol. *27*, 3658-3665.e4.

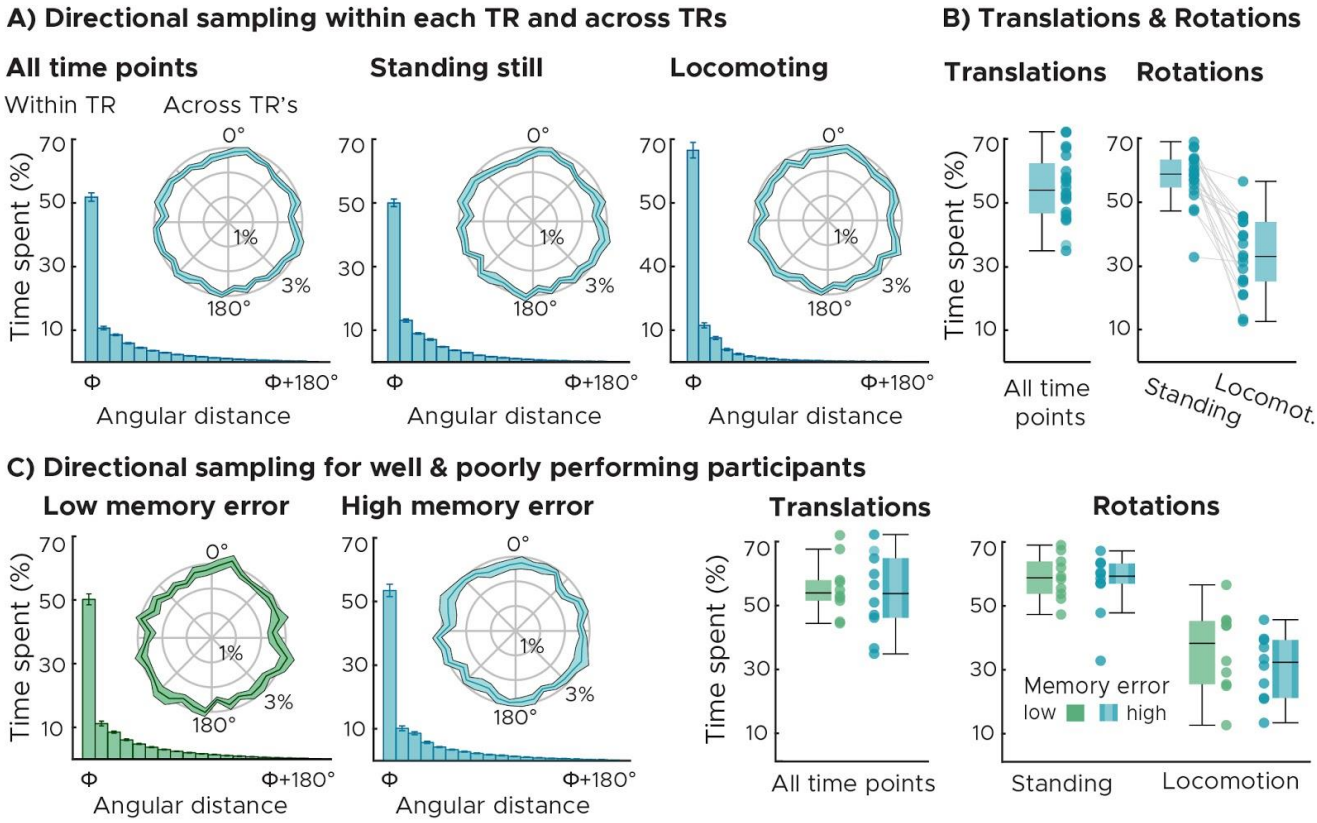
Wirth, S., Baraduc, P., Planté, A., Pinède, S., and Duhamel, J.-R. (2017). Gaze-informed, task-situated representation of space in primate hippocampus during virtual navigation. PLOS Biol. *15*, e2001045.

Witter, M.P., Doan, T.P., Jacobsen, B., Nilssen, E.S., and Ohara, S. (2017). Architecture of the Entorhinal Cortex A Review of Entorhinal Anatomy in Rodents with Some Comparative Notes. Front. Syst. Neurosci. *11*, 1–12.

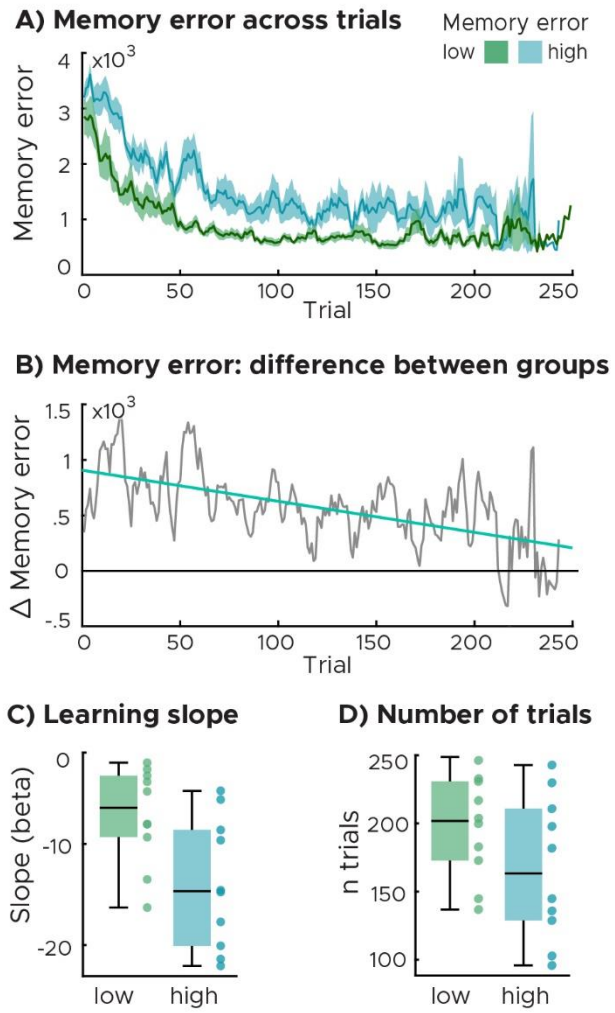
Wolbers, T., and Büchel, C. (2005). Dissociable Retrosplenial and Hippocampal Contributions to Successful Formation of Survey Representations. J. Neurosci. *25*, 3333–3340.

Yarkoni, T., Poldrack, R.A., Nichols, T.E., Van Essen, D.C., and Wager, T.D. (2011). Large-scale automated synthesis of human functional neuroimaging data. Nat. Methods *8*, 665–670.

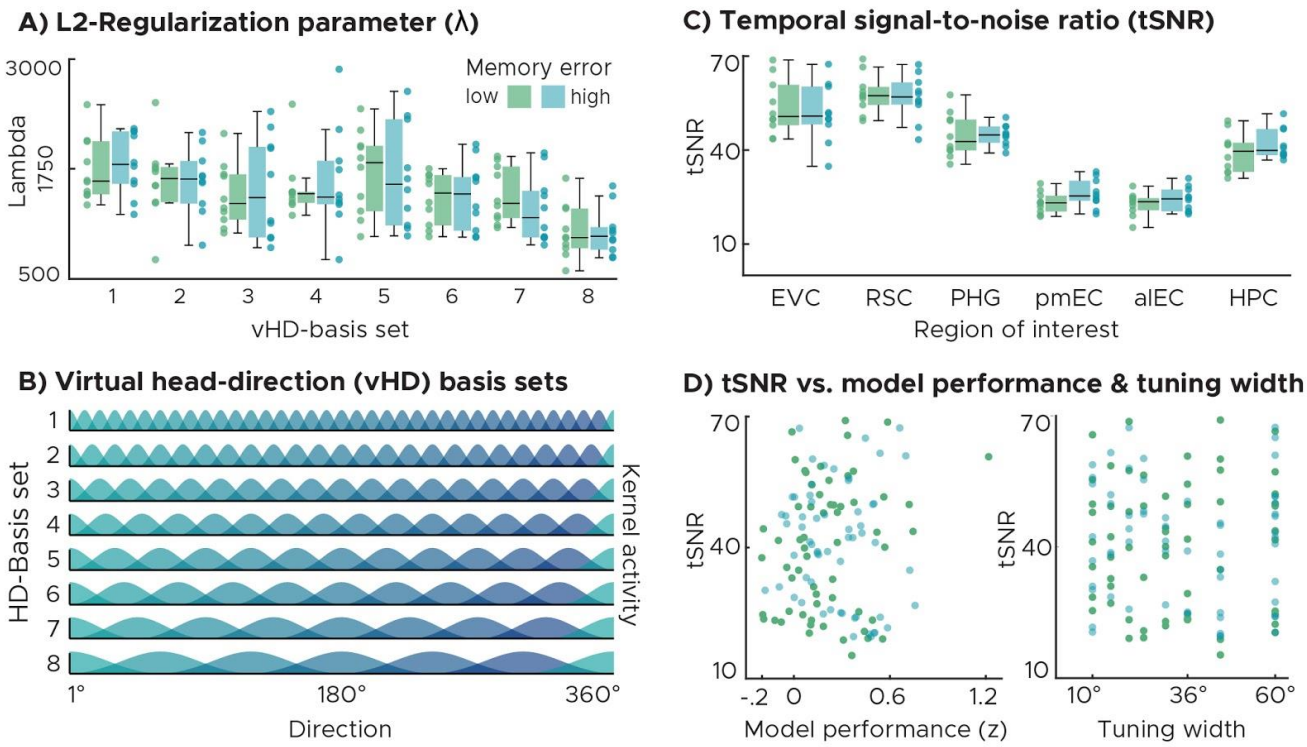
Supplementary material



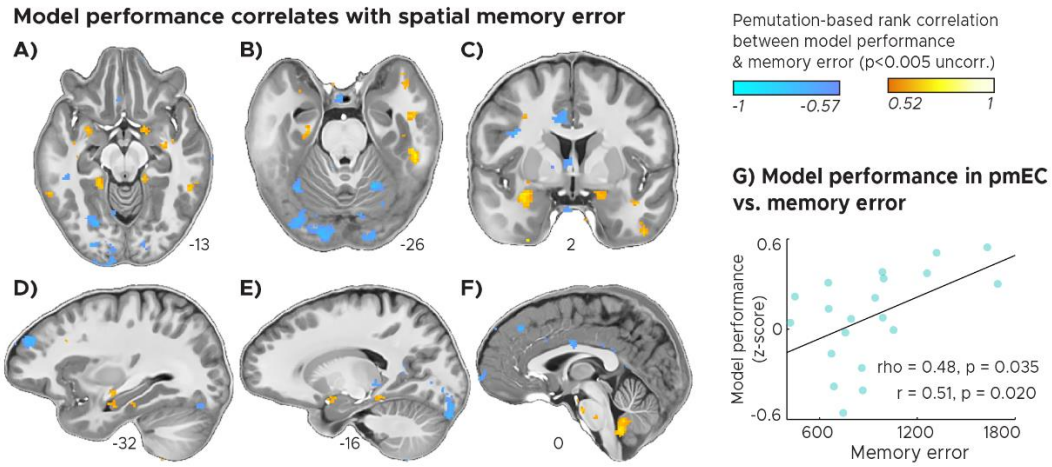
Supplementary figure 1. Navigation behavior (n = 20). A) Directional sampling across all time points (left), stationary periods only (middle) and locomotion periods only (right). Directional sampling expresses the time spent facing into each direction in percent of total experiment time. The bar plots depict the across-participant mean & SEM within each fMRI acquisition (TR). Data were pooled according to the angular distance to the predominant direction (Φ, first bar). Within each TR, participants' faced into one predominant direction for around 50% of the total experimental time. The time spent facing into the predominant direction differed between stationary and locomotion periods (two-tailed permutation-based paired ttest results: t(19) = 6.12, p = 0.0001, k = 10000). Polar plots depict the directional sampling across the entire experiment (line: mean, shaded area: SEM). Across the experiment, sampling was matched across directions, showing that there were no asymmetric spatial cues that biased navigation behavior (rmANOVA results for all time points: F(35, 665) = 0.77 , p = 0.834, while standing still: F(35, 665) = 0.87, p = 0.681 and while locomoting: F(35, 665) = 0.79, p = 0.806). B) Translations and rotations. Participants locomoted around 50% of the total experimental time (left panel). While locomoting, participants spent less time rotating than while standing still (right panel) (two-tailed permutation-based paired ttest results: t(38) = 7.95, p = 0.0001, k = 10000). We plot single-participant data and group-level whisker-and-box plots (center, median; box, 25th to 75th percentiles; whiskers, 1.5 × interquartile range). C) Directional sampling across participant groups. There were no differences in directional sampling in neither of the two participant groups across TRs (rmANOVA results: low-memory-error group: F(35, 315) = 0.87, p = 0.675, high-memory-error group: F(35, 315) = 0.55, p = 0.984). The two participant groups spent equal amount of time locomoting (t(18) = 0.41, p = 0.690), and rotating (while locomoting: t(18) = 0.55, p = 0.625 and while standing still: t(18) = 1.00, p = 0.338) and facing towards one predominant direction within each TR (t(18) = -1.26, p = 0.224, k = 10000).



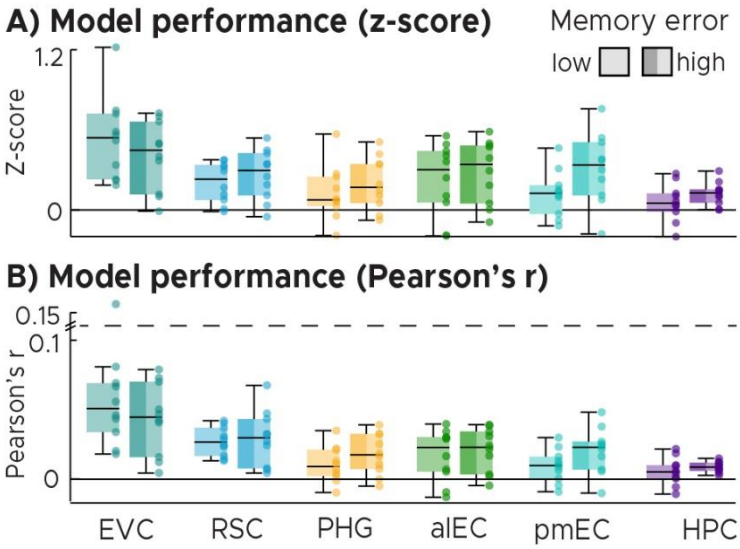
Supplementary figure 2: Memory performance during the object-location memory task. A) Memory error across trials for the two participant groups differing in the median memory error. The memory error is the Euclidean distance between true and remembered object location expressed in virtual vertices. Data were smoothed using a running average kernel of 5 trials. B) Differences in memory error between groups. Zero constitutes no difference. Both groups converge on the same level of performance in the course of the experiment. C) Learning slopes. Regression slopes of a linear line fitted to the raw memory error scores across trials of each participant. Participants with higher median memory error showed steeper slopes (two-tailed permutation-based unpaired ttest results: $t(18) = 2.65$, $p = 0.019$), indicating that the memory error improved faster than in the low-memory-error group. D) Number of trials. Participants with low memory error performed slightly more trials than the ones with high memory errors, the difference between groups was however small ($t(18) = 1.46$, $p = 0.137$). C,D) We plot single-participant data and group-level whisker-and-box plots (center, median; box, 25th to 75th percentiles; whiskers, 1.5 × interquartile range).



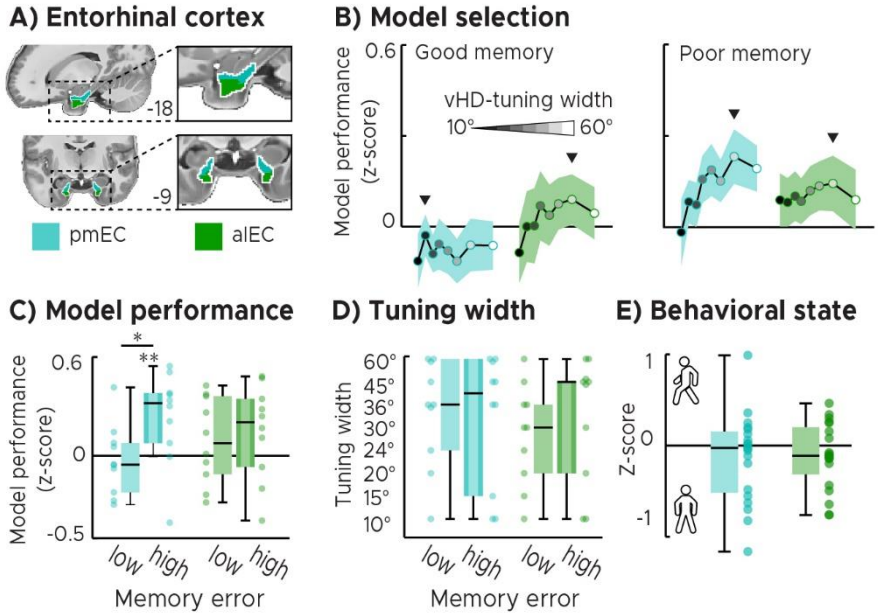
Supplementary figure 3: Model parameters and data quality. A) L2-regularization parameters (λ) for all basis sets and the two participant groups (differing in memory error). Each dot represents λ of a single participant and basis set averaged across all ROIs (Fig. 4). Group-level whisker-boxplots of the same data were added (center, median; box, 25th to 75th percentiles; whiskers, 1.5 × interquartile range). Lambda depended on the basis set (rmANOVA results: $F(7, 126) = 11.80$, $p = 1.7 \times 10^{-11}$) as expected, but not on participant group ($F(1, 18) = 0.0037$, $p = 0.952$). There was no interaction between the two ($F(7,126) = 0.37$, $p = 0.920$). B) Histogram depiction of all basis sets used. Each Gaussian represents one directional kernel covering the full 360° with 1° resolution. Basis sets differed in kernel width and spacing. Resulting regressors were scaled between 0 and 1. C) Temporal signal-to-noise ratio (tSNR) across ROIs and participant groups. We plot group-level whisker-boxplots and single participant data of the average tSNR of each region. There were differences in tSNR across ROIs (rmANOVA results: $F(5, 90) = 258.29$, $p = 9.1 \times 10^{-52}$), but not across participant groups ($F(1, 18) = 0.082$, $p = 0.777$) and there was no interaction between the two ($F(5, 90) = 1.34$, $p = 0.255$). D) Scatter plots for tSNR over model performance (left) and estimated tuning width (right). Dots represent the average tSNR of each region and participant. Neither model performance (Spearman correlation: rho = 0.076, p = 0.409), nor tuning width (rho = -0.059, p = 0.525) correlated with tSNR.



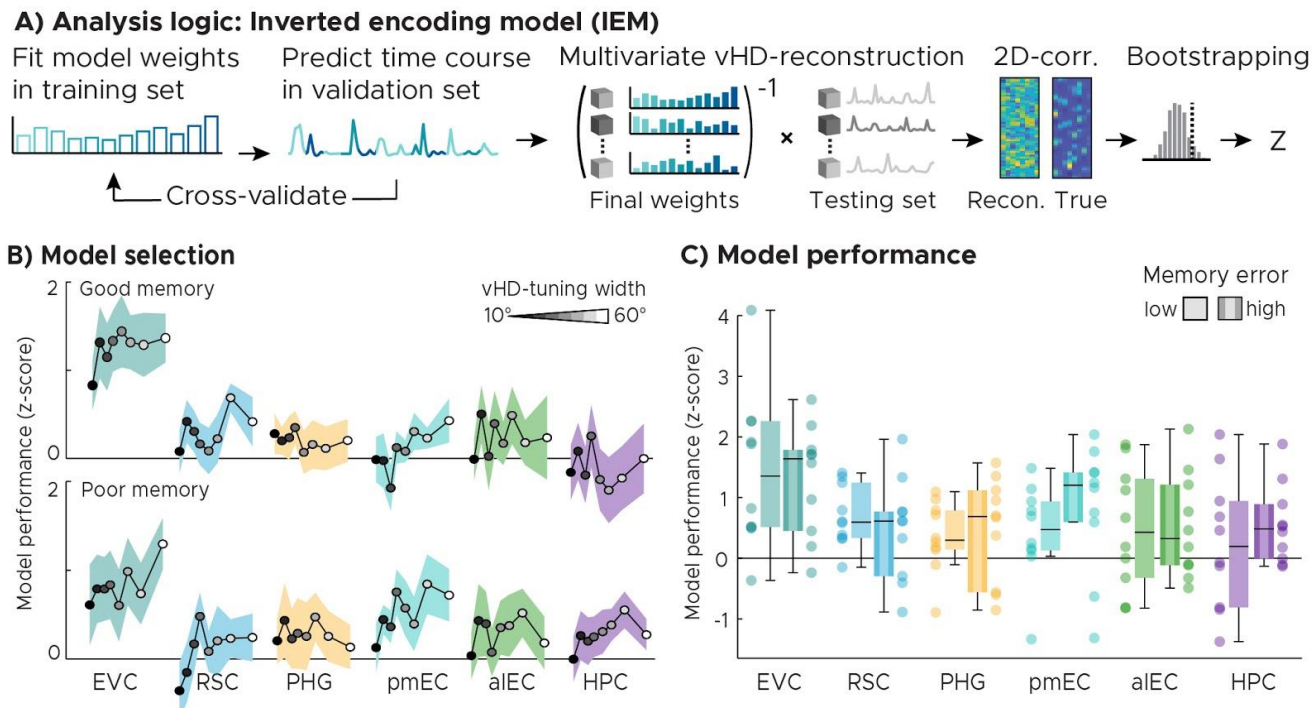
Supplementary figure 4: Across-participant permutation-based rank-correlation between the model performance (see Fig. 3) and across trial median memory error (Fig. 1). A-F) We plot Spearman's correlation coefficients (Rho) thresholded at p<0.005 uncorrected for visualization and overlaid on the group-average T1 template at T1 resolution. We observed a positive correlation between model performance and memory error in the parahippocampal gyrus (A,B,D,E) and anterior medial temporal lobe (A,B,C,D,E) including posterior entorhinal cortex (B) and anterior hippocampus (D), as well as in the ventral cerebellum (F). Negative correlations could be observed in the ventral occipital lobe (A,B,D,E), frontal lobe (D,F), the medial cingulate cortex (C) and the posterior thalamus (E). G) Scatter plot of posteromedial entorhinal cortex (pmEC) model performance (Fig. 5) over across-trial median memory error (Fig. 1). Least-square line as well as permutation-based rank & linear-correlation coefficients were added. There is a correlation between model performance and memory error in pmEC.



Supplementary figure 5: Optimal model performance expressed in Z-scores (A) and Pearson's R (B) selected for each individual participant instead of group level (Fig. 5B,C). We plot single-participant data overlaid on group-level whisker-and-box plots (center, median; box, 25th to 75th percentiles; whiskers, 1.5 × interquartile range) for low and high-memory-error groups. B) Y-axis was cut for visualization (dashed line).



Supplementary figure 6: Comparison between entorhinal cortex subdivisions. A) Regions of interests: posteromedial (pmEC) vs. anterolateral entorhinal cortex (alEC). B) Model selection: We plot the model performance (Z-score) for all basis sets and the two participant groups. The black line and the shaded area represent the mean and SEM across participants. Each dot represents the group-average model performance for one basis set, with darker colors representing narrow kernels and lighter colors representing wider kernels. The following kernel widths were tested: 10°, 15°, 20°, 24°, 30°, 36°, 45°, 60°. For each ROI on group-level, we selected the best performing basis set as the optimal model to be tested. Also see Fig. 5B. C) Optimal model performance selected in B for the two (high- and low-memory-error) participant groups. We plot single participant data and group-level whisker-boxplots (center, median; box, 25th to 75th percentiles; whiskers, 1.5 × interquartile range). We observed directional tuning in pmEC in participants with high memory error and a difference between groups (low memory error: t(9) = -0.42, p = 0.661; high memory error: t(9) = 2.59, p = 0.020, pFDR = 0.040, t(19) = 2.32, p = 0.036). In alEC, neither directional tuning nor the difference between groups could be observed (Low : t(9) = 1.09, p = 0.146; high: t(9) = 1.58, p = 0.075; contrast: t(19) = 0.44, p = 0.655). Also see Fig. 5C. D) Optimal tuning width. Similar to B,C, we plot the tuning width that led to the highest memory performance selected on individual participant level (see Fig. 5D). E) Behavioral state analysis. Model performance during locomotion and stationary periods. Positive values indicate that voxel time courses in the ROI could be better predicted when participants locomoted. We plot the difference in model performance as single participant data and group-level whisker-boxplots (center, median; box, 25th to 75th percentiles; whiskers, 1.5 × interquartile range). There is no effect of locomotion per se on the tuning in pmEC (and alEC which was not directionally tuned in our task).



Supplementary figure 7: Multivariate inverted encoding model (IEM). A) Analysis logic. We estimated voxel-wise weights following the same training procedure used for the forward model (see methods). We then multiplied the Moore-Penrose pseudoinverse of all voxel-wise weights in an ROI (m voxels x k weights)-1 with the multivoxel-pattern (m voxels) at each image acquisition (TR) to obtain the estimated vHD-kernel activities at each TR (k weights). We then used 2D-correlation between the reconstructed kernel activities (k weights x n TR's) and the design matrix of the test run (also k weights x n TR's) to compare reconstructed and observed vHD. Finally, the resulting correlation coefficient was converted to a z-score using the bootstrapped null distribution of each ROI (k = 500 unique shuffles). B) Model selection: We plot the IEM-model performance (Z-score) for all basis sets. The black line and the shaded area represent the mean and SEM across participants. Each dot represents the group-average model performance for one basis set, with darker colors representing narrow kernels and lighter colors representing wider kernels. The following kernel widths were tested: 10°, 15°, 20°, 24°, 30°, 36°, 45°, 60°. For each ROI on group-level, we selected the best performing basis set as the optimal model to be tested. C) Optimal model performance selected in B for the two (high- and low-memory-error) participant groups. We plot single participant data and group-level whisker-boxplots (center, median; box, 25th to 75th percentiles; whiskers, 1.5 × interquartile range).