Title: Origin of hexagonal metric of grid cells in navigating spaces

Bo Zhang[1,2], Yiyuan Zhang[3], and Jia Liu[1*]

[1] Tsinghua Laboratory of Brain and Intelligence & Department of Psychology, Tsinghua University, Beijing, China

[2] Beijing Academy of Artificial Intelligence, Beijing, China

[3] Faculty of Psychology, Beijing Normal University, Beijing, China

*Correspondence shall be addressed to: JL, liujiathu@tsinghua.edu.cn

Abstract

To achieve the computational goal of navigating both natural and mental spaces, the brain adopts a hexagonal metric of grid cells in the entorhinal cortex (ERC) to chart the spaces. However, little is known about how and why the hexagonal metric emerges. Here we designed an object matching task where participants adjusted two parts of object variants to match their prototype in fMRI scanner. Unbeknownst to the participants, the object variants were arranged on a ring centering on the prototype in an object space constructed by the object parts. When the participants navigated from a location in the ring to the center in the space, we observed hexagonal signals in the ERC, and more importantly, a spatial rhythm of 3Hz in the hippocampus (HPC). Accordingly, we built a computational model depicting the spatial rhythm of 3Hz as scaffolds to transform discrete spatial locations into structured states of cognitive map that allows continuous updates of location estimates with each movement made. The formal proof shows that to support the cognitive map in the HPC, the spatial input from the ERC must be in a hexagonal pattern. To further examine the biological plausibility of the model, we built various forms of neural networks in their simplest architecture to transform locations in space

into structured states of cognitive map, and units with hexagonal patterns emerge ubiquitously. In short, with empirical experiments, formal proof, and neural network simulation, our study reveals the origin of the hexagonal metric of grid cells mechanistically.

Introduction

Animals' knowledge on the geometry of external world constructs space in mind to guide their survival (Kant., 2006). The discovery of grid cell (Hafting et al., 2005), whose hexagonal activation pattern seamlessly tessellates space (Bush et al., 2015; Moser et al., 2017), promotes further investigation on the hexagonal metric from natural space to abstract spaces in human including associative memory (Constantinescu., 2016), eye movement (Nau et al., 2018), odor intensity (Bao et al., 2019), reward location (Sosa and Giocomo, 2021) and sequence states (McNamee et al., 2021), revealing the general nature of the hexagonal metric in charting various mental spaces. However, from a mathematic perspective, hexagon is not the only geometric form that can seamlessly tessellate space; at least tetragon and certain irregular pentagons can also chart space seamlessly. Why do animals uniformly adopt the hexagonal metric when navigating spaces?

Numerous theoretical models have been proposed to explore the origin of the hexagonal metric of grid cells. One type of models is called the interference model, which hypothesizes that the hexagonal pattern emerges by the summation of three planar waves with preferred directions 60° apart (Solstad et al., 2006; Burgess et al., 2007; Dordek et al., 2016; Sorscher et al., 2019). Another uses continuous attractor networks to produce the hexagonal pattern by integrating excitatory and inhibitory neurons (Samsonovich and McNaughton, 1997; Guanella et al., 2007;

Burak and Fiete, 2009). Though the hexagonal pattern is successfully generated, these models are usually founded on various biological implementations assumed, such as phase differences (Bush et al., 2014), interaction among neurons (Couey et al., 2013), or characteristics of neural activity (Dordek et al., 2016), rather than on empirical phenomena directly observed when participants performing navigation tasks. Therefore, without proper biological constraints, theoretically there will be infinite number of models could generate the hexagonal pattern, and none may reveal mechanistic nature of grid cells. Here in this study, we first explored biological constraints during the navigation of mental spaces, which were then based upon to build a computational model to deduce the origin of the hexagonal metric of grid cells.

To do this, we constructed an artificial object space and asked participants to match object variants distributed in a ring to the prototype situated in the center (Fig.1). By adjusting the lengths of two object parts independently in fMRI scanner, the participants navigated in the object space from the periphery to the center. By analyzing paths from the variants to the prototype, we expected to identify hexagonal BOLD signals in the ERC and further explored how positional information was encoded and maintained in the HPC. Centering on this empirical observation, we constructed a computational model on the ERC-HPC system that grid cells in the ERC provide spatial input to place cells in the HPC (O'Keefe and Dostrovsky, 1971) to perform the task of path integration (Bush et al., 2014; Bush et al., 2015; Evans et al., 2016). Then, we presented a formal proof to demonstrate how and why the hexagonal pattern emerges generically when the ERC-HPC system optimally solves the computational problem of path integration. Finally, to test the biologically plausibility of the model, we illustrated the emergence of grid-cell-like units in various artificial neural networks (ANN) when their

connections were tuned to optimally encode positions in space into structured states of cognitive map.

Results

Construction of an artificial object space

A novel 3D object, named greeble (Gauthier and Tarr, 1997) (Fig.1a, Top), was used to construct an artificial object space, where a location in the space (Fig.1b) indicated a greeble with unique lengths of two parts, “Loogit” and “Vacso”, that formed the axes of the space. The greeble in the center was the prototype, serving as the target in an object matching task. In the task, a greeble variant from the ring (Fig.1b, red points in the ring) was presented, and participants were instructed to adjust the length of Loogit and Vasco independently with button press so that the variant matched the prototype stored in memory as accurately as possible (Fig.1a, Bottom). A path for one trial in the object space was determined by the location of the variant in the ring and the location where the participant stopped (Fig.1c). To ensure the space was fully covered, we evenly chose variants in the ring to obtain the paths ranging from 1° to 360° (see Methods for details). Note that in the task, no participant was aware of the existence of the object space or had a subjective feeling of navigating in space. The behavioral results are shown in Fig.S1.

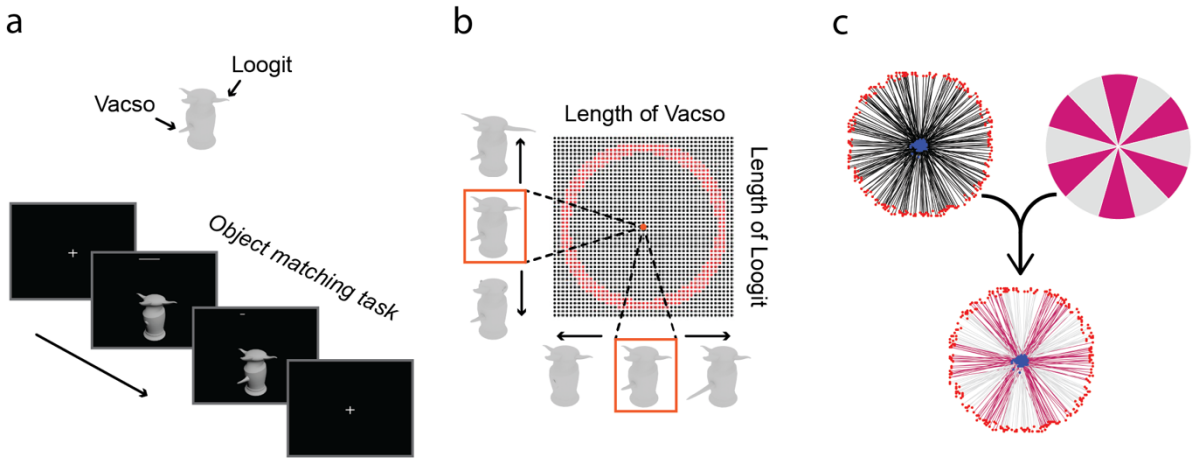


Figure 1. Experimental design and object space. (a) Top: A greeble exemplar and its two parts: "Loogit" and "Vacso". Bottom: In the scanner, participants continuously adjusted the length of Loogit and Vacso to match the prototype that was learnt before the experiment by pressing buttons within a 10-s time limit. (b) The artificial object space. Each red point in the ring represents a variant greeble in a trial (i.e., start location), the red point in the center indicates the prototype greeble (i.e., goal location). Each black point in the space represents an intermediate greeble that may connect the variant to the prototype to construct a path. (c) The paths of a typical participant reconstructed from the button press during the experiment (top left), the schema for the pattern of six-peaked sinusoidal modulator (top right), and the overlapped pattern showing evenly distributed paths aligned with grid axes.

Hexagonal activity in the ERC

First, we examined whether the hexagonal metric was recruited in charting the object space when the participants performed the object matching task by localizing the six-peaked sinusoidal activity in the ERC (Fig.1c, see Methods for details). Significant uniformity of hexagonal orientations within the ERC was observed in all participants (Fig.S2, $p < 0.05$, Rayleigh test of uniformity, Bonferroni corrected). As expected, hexagonal signals were successfully revealed in the ERC (right hemisphere, $p < 0.01$, initial threshold; $p < 0.05$, small-volume cluster-corrected (SVC) for multiple comparisons) (Fig.2a). In contrast, no significant rhythmic activity of 5-

round (72° periodicity) or 7-round (51.4° periodicity) modulations was found in the ERC. An ROI analysis on the ERC confirmed that robust rhythmic activity was only revealed by 6-round modulation (Fig.2b, t(9) = 3.89, p = 0.003), but not for 5-round or 7-round modulations (Fig.2b, 5-round: t(9) < 1, 7-round: t(9) < 1) in the ERC. Those results suggested that grid cells in the ERC were activated when the participants implicitly navigated in the object space to find path from a variant greeble to the prototype greeble in the object matching task.

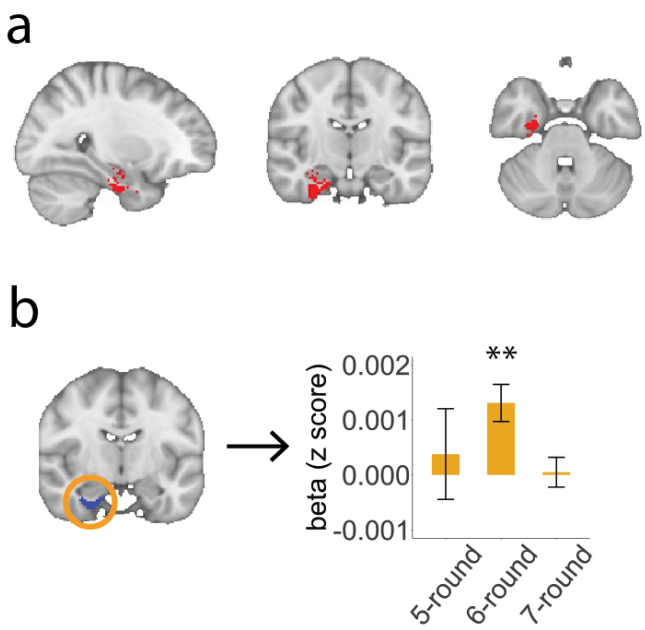


Figure 2. Localization of hexagonal activity when participants navigated in the object space. (a) Hexagonal activity was observed in the right ERC (p < 0.01, initial threshold; p < 0.05, SVC for multiple comparisons, peak MNI coordinate: 19, -15, -27; peak t = 4.58). (b) An ROI analysis illustrated the BOLD signals of 5-round, 6-round, and 7-fold modulations using the mask of right ERC (t(9) = 3.89, p = 0.003), which was manually delineated from the MNI T1 brain template packaged by FSL. **: p < 0.01. Error bars indicate SEM.

Spatial rhythm of 3Hz in HPC CA1

Having confirmed the involvement of grid cells in the ERC when participants navigated in the object space, we further examined how positional information in the space was encoded and maintained in the HPC. Inspired by the pioneer work (Hubel and Wiesel, 1959; Hubel and Wiesel, 1962) that rats' ganglion cells in the striate cortex encode the location and orientation of a moving light spot with spatial rhythm of neural activity (see also Bieri et al., 2014; Fiebelkorn et al., 2018) and the fact that spatial rhythm, as a vector quantity, naturally provides directionality necessary for representing vector relationship among spatial locations (Bush et al., 2015; Stemmler et al., 2015), here we explored potential spatial rhythms of neural activity in the HPC during navigation (Fig.3a). Unlike previous studies with low spatial resolution adapted to the 60° spatial periodicity of grid cell (e.g., Doeller et al., 2010; Constantinescu et al., 2016), our design captured almost all path orientations ranging from 1° to 360° with the precision of 1° (Fig. 1c) that resulted in sufficient number of samples for spectrum analyses in spatial domain (De Valois et al., 1982). Specifically, the spatial orientations of paths were first down-sampled to 15° bins, resulting in a vector of resampled spatial orientations containing 24 samples (e.g., 0°, 15°, and 30°). After the activity maps of spatial orientations were estimated and sorted in an ascending order, Fast Discrete Fourier Transform (FFT) was performed on the BOLD signal series along the spatial orientations to extract the spatial spectrum from 0hz to 11hz (see Methods for detail). The result revealed a significant cluster of FFT magnitude at 3hz in the anterior portion of the HPC (right hemisphere, $p < 0.05$, initial threshold; $p < 0.05$, SVC for multiple comparisons)(Fig.3b). Careful examination on the anatomic location of the spatial rhythm with the anatomical subfields of the HPC (Iglesias et al., 2015) revealed that the strongest FFT magnitude at 3hz presented in the CA1 of the HPC. Previous studies have shown that the CA1 is anatomically connected to the ERC (Bush et al., 2014; van Groen et al., 2003),

implying the role of the HPC in signal back-projection to the ERC for hexagonal metric formation. Supporting this intuition, although no significant spatial rhythm was found in the ERC, there was significant phase coupling between the ERC and HPC only at 3hz (Fig. 3c, r = 0.79, p = 0.005). No significant phase coupling was found either in other frequencies tested (*p*s > 0.05) or between the HPC and the lateral occipital cortex (LOC), a region specific for object processing (Grill-Spector, 2003) (*p* > 0.05).

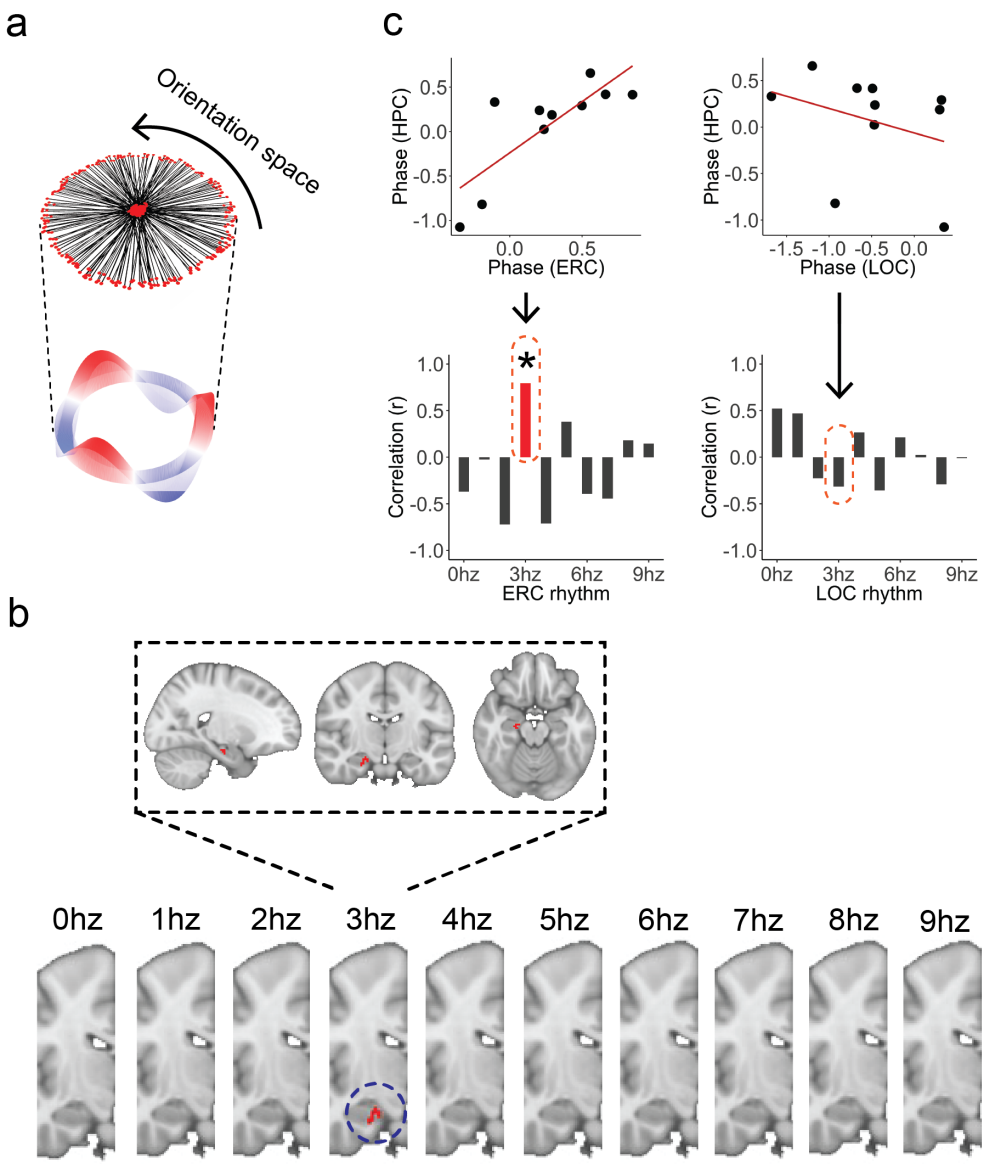


Figure 3. The spatial rhythmic activity of 3Hz in the HPC. (a) A schema illustrating 3hz rhythmic activity in the HPC. Top: an exemplar of paths' spatial orientations from an individual participant; Bottom: a top view of the response field of 3Hz spatial frequency. (b) A spectrum analysis of BOLD signals in spatial domain. A significant cluster of FFT magnitude at 3hz was observed in the right HPC ($p < 0.05$, initial threshold; $p < 0.05$, SVC for multiple comparisons; Peak MNI coordinate: 20, -14, -18). (c) Phase coupling between the HPC and the ERC. Significant correlation of phases from 3hz BOLD signal series across participants was found between the HPC and ERC ($r = 0.79$, $p = 0.005$). In contrast, no significant correlation was revealed between the HPC and the LOC. *: $p < 0.01$.

Replication

Because both the involvement of the ERC in an object matching task and the spatial rhythmic neural activity of 3Hz in the HPC have not been reported in either neurophysiological or fMRI studies before, here we examined the reliability of these results with a completely new group of participants scanned at a different imaging center (i.e., the replication group). The experimental setup, data preprocessing and analyses remained identical with those for the experimental group. In short, we replicated all major findings reported in the experimental group (Fig.S3). That is, we found the involvement of grid cells in the ERC (Fig.S3c, $p < 0.01$, initial threshold; $p < 0.05$, SVC for multiple comparison), and the spatial rhythm of 3Hz in the CA1 of right HPC (Fig.S3d, $p < 0.05$, initial threshold; $p < 0.05$, SVC for multiple comparison). In addition, only at 3Hz did the HPC phase-couple with the ERC (fig. S3e, $r = 0.64$, $p = 0.04$). That is, the novel findings of the involvement of grid cells in the ERC and 3hz spatial rhythm in the HPC are reliable.

Biologically-inspired computational model on path integration

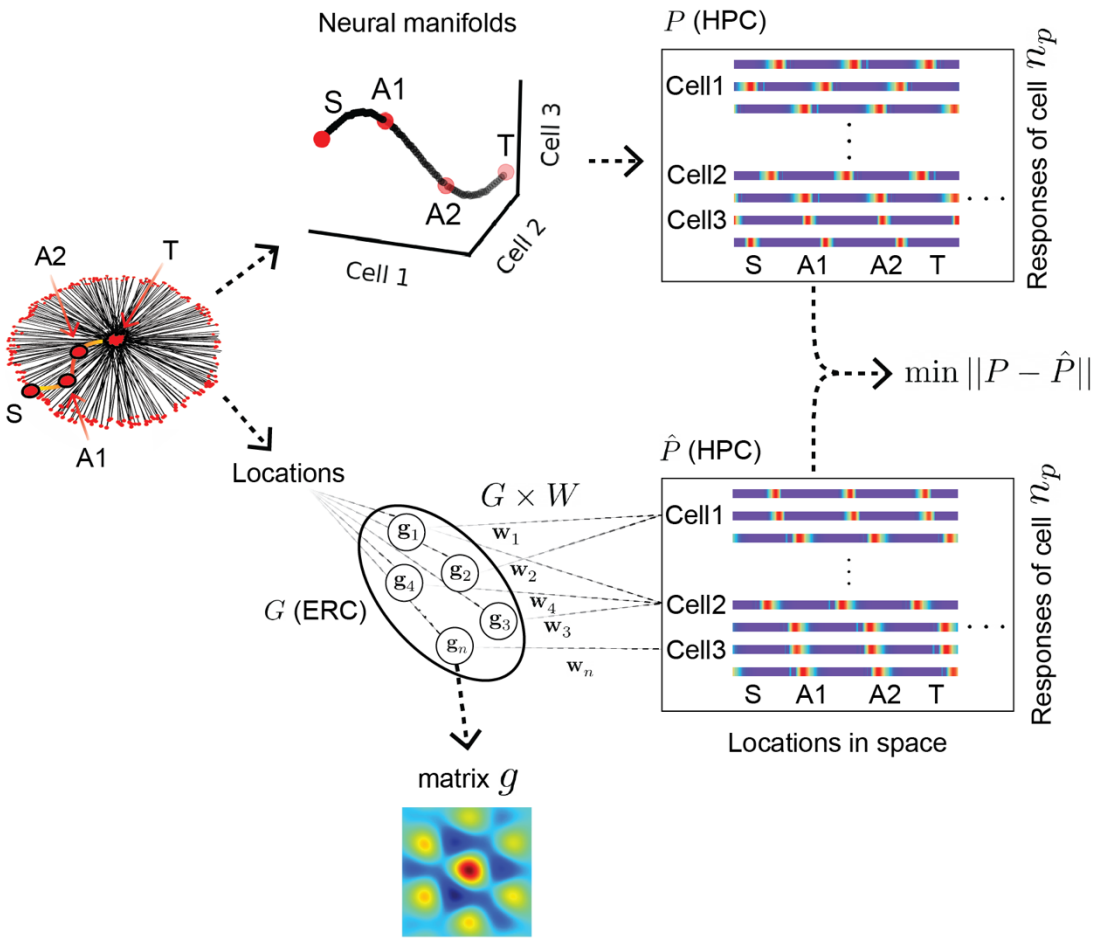


Figure 4. Model on path integration. Top row, path integration is achieved through neural manifolds constructed by place cells in the HPC whose conjunctive neural responses of place field and spatial rhythm, matrix $P(HPC)$, encode the start (S), intermediate (A1 & A2), and goal (T) locations of the space. Colors in matrix P indicate magnitudes of neural activity. Bottom row, biological realization of path integration is through the ERC-HPC system where grid cells in the ERC, $G(ERC)$, provide spatial input (matrix g) to place cells in the HPC to construct matrix $\hat{P}(HPC)$. Because matrix P is known in the model, the minimization of distance between matrix P and matrix $\hat{P}$, $min\|P - \hat{P}\|$, will deduce matrix g, which is neural activity pattern of grid cells. The formal proof demonstrates that matrix g is in a hexagonal pattern.

Centering on the empirical observation of the hexagonal pattern in the ERC and the spatial rhythm of 3Hz in the HPC, we built a computational model to depict how path integration is

achieved, where place cells in the HPC represent single locations in space and grid cells in the ERC provide a spatial metric system not available from place cells to form a structured map (see also Dordek et al., 2016; Sorscher et al., 2019; Sorscher et al., 2020). As shown in Fig.4 Top, a path from a variant (i.e., the start, S) to the prototype (i.e., the target, T) is formed by connecting a series of intermediate variants (e.g., A1 and A2). Each object, including S and T, is a location in space, which can be represented by a neural manifold coded by conjunctive neural responses of place field and spatial neural rhythm in the HPC.

Specifically, the conjunctive response of a place cell is modeled by two parts, place field $\Gamma(\mathrm{s},\mathrm{c})$ of place cells and 3Hz spatial rhythm $\psi(s,c)$ in the HPC observed in this study.

$$f(s,c) = \Gamma(s,c) * \psi(s,c) \tag{1}$$

The place field $\Gamma(\mathrm{s},\mathrm{c})$ is modelled by a Mexican hat function $\Gamma(\mathrm{s},\mathrm{c})$ (Amari, 1977; McNaughton et al., 2006), where s represents the location (s_x,s_y) in a 2-dimentional space χ, and c represents the location (c_x,c_y) of the peak response of the Mexican hat function. $\Gamma(\mathrm{s},\mathrm{c})$ is expressed in equation (2), where $\alpha > \beta$ and $\sigma_1 < \sigma_2$.

$$\Gamma(s,c) = \alpha e^{-\frac{\|s-c\|^2}{2\sigma_1^2}} - \beta e^{-\frac{\|s-c\|^2}{2\sigma_2^2}} \tag{2}$$

The spatial rhythm $\psi(\mathrm{s},\mathrm{c})$ is expressed in equation (3), where the parameter *3* represents spatial frequency v at 3Hz, and the phase angle $\theta(s,c)$ is expressed in equation (4).

$$\psi(s,c) = \cos(3\theta(s,c))$$

(3)

$$\theta(s,c) = \begin{cases} \arctan\left(\frac{s_2 - c_2}{s_1 - c_1}\right), & s_1 - c_1 > 0 \\ \arctan\left(\frac{s_2 - c_2}{s_1 - c_1}\right) + \pi, & s_1 - c_1 < 0 \end{cases} \tag{4}$$

That is, the conjunctive response of a place cell, $f(s,c)$, is the Hadamard product of integrating its place field Γ with 3Hz spatial frequency ψ. That is, the place field of a place cell to a location in space is vectorized and convolved with a spatial frequency v at 3hz. This is critical, because locations, as zero-dimensional points in space (i.e., scalar quantity) that are randomly allocated to place cells (Rich et al., 2014), are connected by scale-free spatial rhythm as structured neural manifolds. Therefore, the path from S to T follows the trajectory of neural manifolds smoothly evolving along time that can continuously update location estimates of variants (e.g., A1 and A2) in the object space when the participants continuously adjusted the length of object parts (Fig.4, Top).

The ensemble of the conjunctive responses from all HPC place cells is expressed as matrix P (Fig.4, Top), where $P \in \mathbb{R}^{n_x \times n_y}$. In the matrix, each row represents the conjunctive response of a place cell n_p in response to all locations in space, and each column is the neural manifold for a location in space constructed by all place cells in the HPC. That is, matrix P effectively indicates which neural manifolds are neighbors and therefore encapsulates a 1-dimensional manifold topology for a path in the 2-dimentional object space. Because matrix P contains neural manifolds for all possible paths from S to T, it is the structured knowledge on the space that it represents. In other words, matrix P is the formal expression of cognitive map (Tolman, 1948).

Biologically, matrix P in the HPC comes from spatial input of grid cells in the upstream ERC (Fig.4, Bottom), which is expressed as the neural activity of grid cells G multiplied by connection weights W between the ERC and HPC.

$$\hat{P} = GW, \ \ where \ \ G \in \mathbb{R}^{n_x \times n_g}, \mathrm{W} \in \mathbb{R}^{n_p \times n_g} \tag{5}$$

In equation (5), matrix $\hat{P}$ represents the biologically-realized cognitive map with the support of spatial input from grid cells. Given that matrix P is known, the neural activity pattern of G can be derived by solving the optimization problem $\|P - \hat{P}\|$ to minimize the distance $\mathfrak{D}$, as shown in equation (6):

$$\mathfrak{D}(G, W) = \left\|P - \hat{P}\right\|_F^2, \ \ where \ \ \hat{P} = GW \tag{6}$$

Based on the formula derivation suggested by previous studies (Sorscher et al., 2019; Sorscher et al., 2020), we solved this optimization problem. First, to simplify the calculation of minimizing the distance $\mathfrak{D}$, the columns of G are constrained to be orthogonal by choosing the invertible matrix of Z and Z^{-1} for G and W, respectively, and performing linear transformation $Z: G \rightarrow GZ$ and $Z^{-1}: W \rightarrow Z^{-1}W$; we then obtained the equation (7):

$$\min_{G} \|P - GG^T P\|_F^2, \ \ s.t. \ \ G^T G = I \tag{7}$$

Thus, the optimization problem can be solved by Lagrangian multiplier as shown in the equation (8).

$$\mathcal{L}(G, \lambda) = \mathrm{Tr}[G^T \Sigma G + \lambda(I - G^T G)], \ \ where \ \ \Sigma = PP^T \tag{8}$$

Equation (8) can be further simplified by considering a single cell g instead of the cell population G in the ERC (i.e., n_g equals to 1) to obtain the equation (9) after replacing the matrix G with vector g:

$$\mathcal{L}(g, \lambda) \ = g^T \Sigma g + \lambda(1 - g^T g) \tag{9}$$

Next, the gradient descent algorithm is used to iteratively reach the local minima of $\mathcal{L}$, where the derivative of g with respect to time is zero, as shown in equation (10) and (11).

$$\frac{dg}{dt} = \Sigma g^* - \lambda g^* = 0 \tag{10}$$

$$\Sigma g^* = \lambda g^* \tag{11}$$

Given that $\mathcal{L}(\mathrm{g}, \lambda)$ is a concave function of g, a fixed point g^* (i.e., stationary g) can be revealed by a perturbation method. Specifically, we put g in a neighborhood near the fixed point g^*: $g = g^* + \epsilon$, where ϵ is a small quantity. Equation (12) is then derived from equation (10) by plugging g:

$$\frac{d\epsilon}{dt} = -\lambda\epsilon + \Sigma\,\epsilon \tag{12}$$

Let $\epsilon = \sum_i c_i\xi_i$, where $c_i \in \mathbb{R}$ and is the projection of ϵ on ξ_i, which is the ith eigenvector of variance-covariance matrix Σ, as shown in equation (13)

$$\sum_i \frac{dc_i}{dt}\xi_i = -\lambda \sum_i c_i\xi_i + \sum_i c_i\lambda_i\xi_i \tag{13}$$

Given that matrix Σ is symmetric with its eigenvectors orthogonal, the eigenvectors are normalized to unit length as $\xi_i^T\xi_j = \delta_{ij}$; then, equation (14) and (15) are derived

$$\frac{dc_i}{dt} = -\lambda c_i + \lambda_i c_i \tag{14}$$

$$c_i \propto e^{(\lambda_i - \lambda)t} \tag{15}$$

Therefore, c_i, as the projection of ϵ on ξ_i, is stable only when $\lambda = max\{\lambda_i\}$. That is, the stationary g^* is the eigenvector corresponding to the maximum eigenvalue of variance-covariance matrix Σ as shown in equation (11). In this way, the optimization problem of $\|P - \hat{P}\|$ is solved.

To demonstrate what g^* looks like, we performed a simulation of 10,000 place cells in the HPC with a field size of 2% relative to the size of space (Muller et al., 1987; see Methods for details) to encode 10,000 random locations in space. By performing the optimization stated above, we

obtained g^* that is the maximum eigenvalue of matrix Σ. The g^* showed a regular hexagonal pattern when it was projected onto the space (Fig.4, matrix g). That is, the hexagonal pattern of grid cells in the ERC results from the requirement of constructing cognitive map in the HPC in representing spaces, where in this experiment the participants successfully matched a variant greeble to the prototype greeble by adjusting the length of the object parts.

Biological plausibility of the model

The hexagonal pattern of grid cells in the ERC derived from the computational model is purely a formal proof, and therefore it is unknown whether such computational model is biological plausible. To address this issue, here we used ANNs as a simulator of biological neural network of the ERC and examined whether the grid-cell-like units could automatically emerge in the ANNs when they were trained to generate cognitive map (i.e., matrix P) in the HPC with locations in space as an input (Fig. 5a). To test the general applicability of the computational model, three typical ANNs were employed, including convolutional neural network (CNN), recurrent neural network (RNN), and multiple layer perceptron (MLP). Critically, the ANNs were built in the simplest form of both architecture and functions. That is, only one hidden-layer was used to simulate grid cells in the ERC, momentum was avoided by loss function, and no activity function was applied.

As shown in Fig.5d & Fig.S4, grid-cell-like units were automatically emerged when the ANNs were trained to generate cognitive map in the HPC, matrix P (see Method for detail). Specifically, in addition to search for units with grid-like responses as evidence of successfully generating grid cells as previous studies did, here we adopted much stricter criteria of evaluating

all units in the hidden-layer as a whole. This is important, because a small number of grid-cell-like units may not be sufficient to generate hexagonal metric system. First, 94%, 67%, and 76% units, out of the 100 hidden units in CNN, RNN, and MLP, respectively, showed grid-like activity with their gridscore exceeding the threshold derived from the field shuffling procedure (Fig.5c, ps < 0.05, permutation test, corrected for multiple comparison). Importantly, for all ANNs, the mean gridscore of 100 hidden units exceeded the chance level threshold (mean gridscore = 0.81, 0.78, 0.80, threshold = 0.65, 0.72, 0.79 for CNN, RNN, and MLP, respectively, ps < 0.05, permutation test, corrected for multiple comparison) (Fig.5b). Note that in simulation, not all units in the hidden layer showed the hexagonal pattern, indicating that units in the neural networks showed both regular hexagonal and random activity patterns. This phenomenon is similar to heterogeneous activity patterns observed in the ERC (Krupic et al., 2012; Domnisoru et al., 2013) and supports the theoretical analysis that precise symmetry is not necessary for accurate population-based spatial coding (Stemmler et al., 2015). In short, our model is biological plausible, which unlikely relies on ANNs' particular architecture or functions.

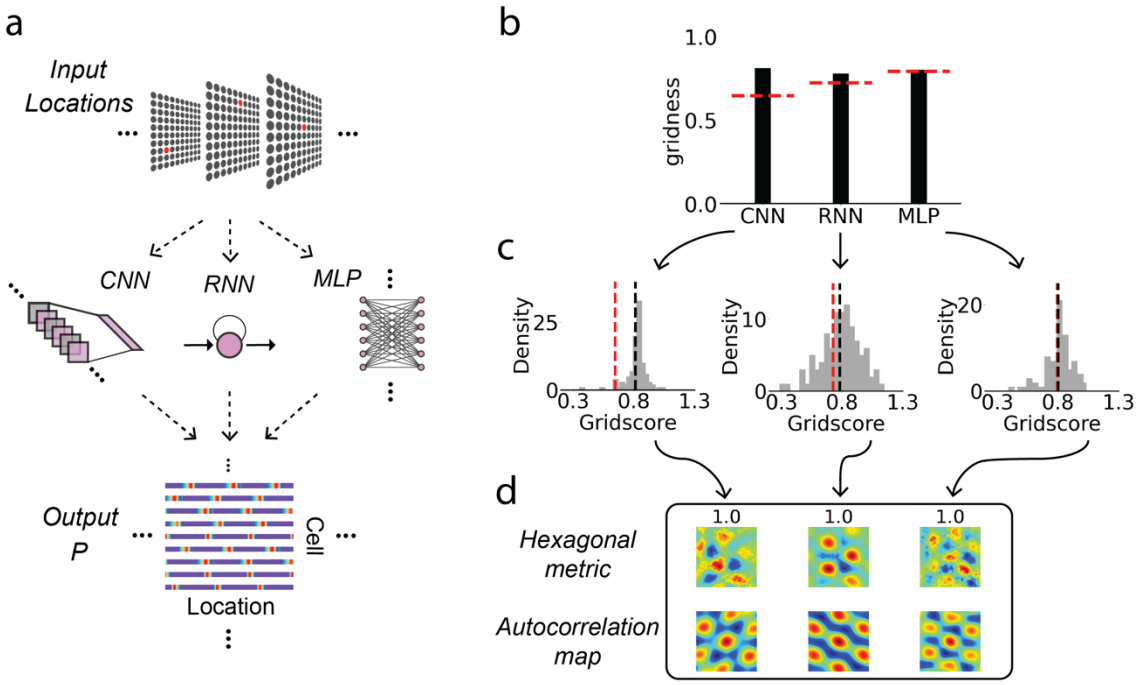


Figure 5. Emergence of grid-cell-like units when training ANNs to generate cognitive map. (a) One-layer CNN, RNN, and MLP were built with locations in space as an input and cognitive map, matrix *P*, as output. (b) Mean gridscore of all hidden units in ANNs. Dash lines represent the threshold of gridness derived from the field shuffling procedure. (c) The distribution of the gridscores of all 100 hidden units. Red dash lines indicate 95th percentile of the permutation distribution, and black dash lines indicate mean gridscore shown in (b). (d) Exemplar hexagonal patterns of grid-cell-like units (the 1st row), which were plotted by overlapping units' activity on space, and the number indicates the gridscore of the corresponding hexagonal pattern. The 2nd row represents the autocorrelation map of the hexagonal patterns.

In both computational model and neural network simulation, the empirical observation of 3Hz rhythmic activity in the HPC played a critical role in forming the hexagonal metric system. Here we further asked whether the hexagonal metric system could emerge at spatial frequencies other than 3Hz. To do so, units' gridness in the hidden layer of the ANNs was calculated when spatial rhythms activity of 0 – 9 Hz (except 3Hz) were applied, respectively. As expected, none was capable of producing the hexagonal metric system, as the mean gridscore of all 100 hidden units

did not exceed the significance threshold (*p*s >0.05, permutation test, corrected for multiple comparison) (Fig.6a). Note that only at the spatial rhythm of 9Hz were there a small number of units whose gridness score was higher than the threshold (1, 12, and 3 units in CNN, RNN, and MLP, respectively). Close inspection on the activity pattern of these units shows that they are distinctively different from the hexagonal pattern of grid cells observed in the ERC (Fig.6b). In short, those results suggest that the spatial rhythm of 3Hz observed in the HPC is necessary to form the hexagonal pattern of grid cells in the ERC.

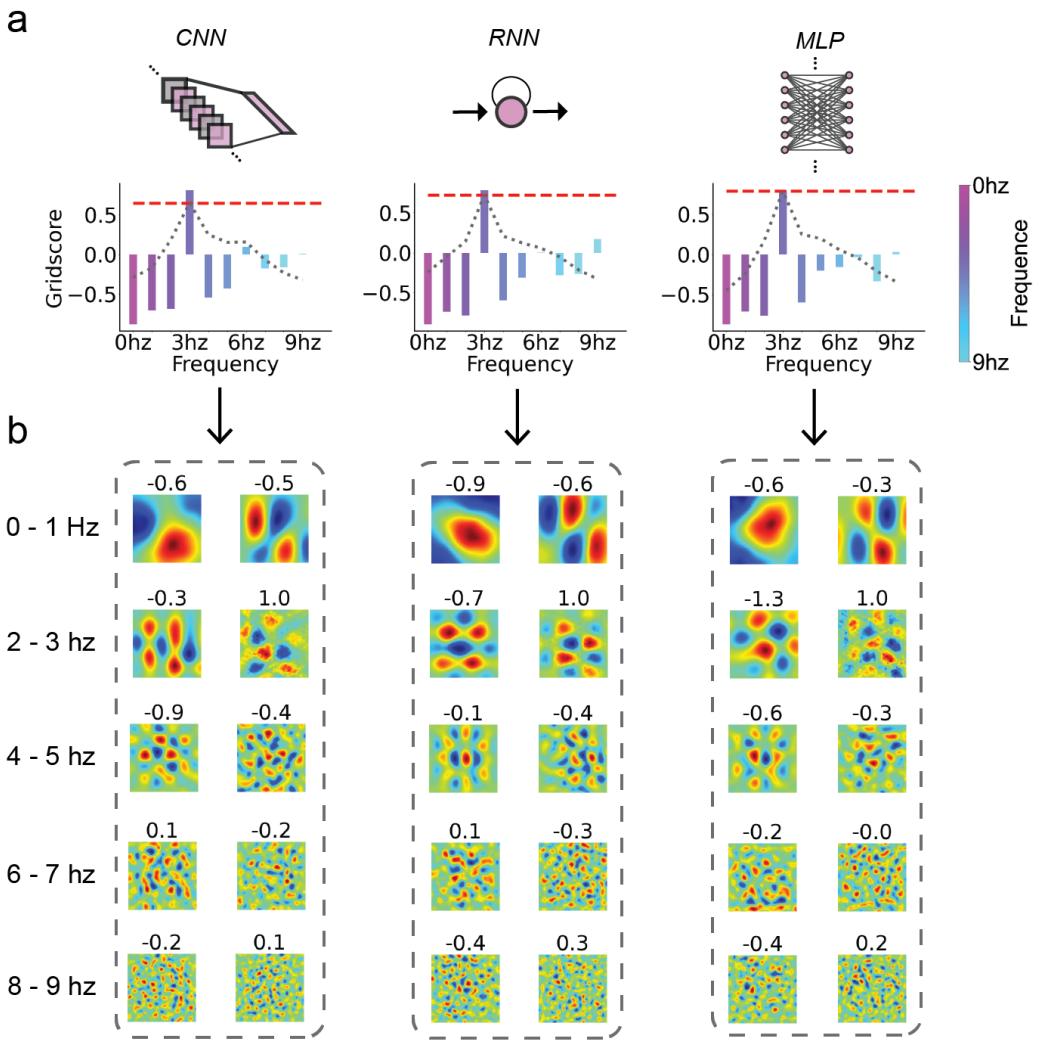


Figure 6. Hexagonal patterns from ANNs with spatial rhythms ranging from 0 – 9H. (a) Mean gridscores of all 100 hidden units for each spatial rhythm tested. Gray and red dash lines indicate uncorrected and corrected thresholds of gridness, respectively. (b) Exemplar activity patterns of units in ANNs for all spatial frequencies tested.

Discussion

Since the discovery of grid cells in a variety of animals and human (Hafting et al., 2005; Yartsev et al. 2013; Killian et al. 2012; Doeller et al. 2010) and its ubiquitous role in navigating numerous mental spaces (Hafting et al., 2005; Constantinescu., 2016; Nau et al., 2018; Bao et al., 2019; Sosa and Giocomo, 2021; McNamee et al., 2021), great attention has been drawn to the mechanistic question of how and why the hexagonal activity pattern of grid cells is formed. Here we addressed this question at three levels emphasized by Marr (1982) to pry open the mystery of how the mental world is charted. At the highest level of computation, we built a model that links the ERC and the HPC to optimally solve the computational problem of path integration that transforms discrete locations in space into structured states of cognitive map. At the middle level of algorithm, the model adopted the spatial rhythm of 3Hz observed in the HPC to encode positions in a path as 1-dimenstional topology of neural manifolds, and the formal proof demonstrates that to achieve the optimal solution, the activity pattern of grid cells in the ERC must be in a regular hexagonal pattern. Finally, when the algorithm was implemented in various forms of ANNs, grid-cell-like units emerged generically, showing the biological plausibility of the model. In sum, with the empirical experiment, formal proof, and neural network simulation, our study reveals the origin of the hexagonal metric system of grid cells in charting mental spaces.

Previous studies have revealed that the CA1 of the HPC mainly receives input from the ERC via temporoammonic projections (Witter and Amaral, 1991; van Groen et al., 2003), and the primary determinant of place cells and grid cells is path integration that plays a fundamental role in spatial coding in the ERC-HPC system (Solstad et al., 2006; Bush et al., 2014; Bush et al., 2015; Evans et al., 2016). Based on these findings, we built a computational model to establish a causal link that the ERC provides a metric system for the path integration in the HPC from a start position (e.g., a variant greeble in this study) to the target position (e.g., the prototype greeble) in space. The feature that distinguishes our model from others is the spatial rhythm of 3Hz observed in the HPC that integrates receptive fields of place cells possibly through a neural binding process (Von der Malsburg, 1999). This feature is critical, because each location in a path corresponds to a cognitive state that is confined by the geometry of neural manifold constructed by conjunctive neural responses of place cells and spatial rhythm in the HPC. In this way, locations, as zero-dimensional and non-continuous points in space, are connected by scale-free spatial rhythm as structured cognitive states (i.e., matrix P in the model). Therefore, the structured cognitive states of locations allow continuous updates of location estimates with each movement made for path integration. More importantly, this also allows for correct predictions of new paths when approaching from an entirely new direction because the structured states provide knowledge on space. In short, the spatial rhythm observed in the HPC likely provides scaffolds for structured knowledge of the relationship among different cognitive states and thus constructs cognitive map in the mind (Bellmund et al., 2018; Behrens et al., 2018). To support this explicitly-embedded cognitive map in the HPC, our formal proof shows that the spatial input provided by the ERC must be a regular hexagonal pattern. In addition, our model is also biological plausible, as when implementing the model in a variety of ANNs with the simplest

architecture and functions, units with the hexagonal pattern emerge ubiquitously without any additional prerequisites or assumptions.

A remaining question is why there is a spatial rhythm of 3Hz in the HPC, as shown in the fMRI experiment the spatial rhythm of 3Hz was the only reliable spatial frequency identified in the HPC. A possibly explanation comes from neural network simulation that with spatial frequencies other than 3Hz, the activity of the units in the ANNs was not a hexagonal pattern. Therefore, the metric system constructed by the units with spatial frequencies other than 3Hz is not able to seamlessly tessellate the space, which likely results in an imperfect cognitive map. Because grid cells are evolutionarily conservative across a variety of mammals, including rats (Hafting et al., 2005), mice (Fyhn et al., 2008), bats (Yartsev et al. 2013), monkeys (Killian et al. 2012), and humans (Doeller et al. 2010), it is possible that the dominance of 3Hz in the HPC emerged from evolution. That is, at the very beginning, the HPC might have used full-spectrum spatial rhythms to represent animal's self-motion cues, such as optical flow, vestibular signals, and motor efference that contain rich spatial information (Moser et al., 2017). However, when an individual's HPC had happened to possess more of 3Hz rhythmic activity, a metric system with more regular hexagonal pattern in the ERC would have likely been formed. Accordingly, this individual might have possessed better evolutional advantage in efficiently exploring the external world. Neural network simulation to "re-run" evolution in navigation may help addressing this question.

The world we live is highly structured, as well as artefacts we manufactured, such as objects of daily appliance (Huang et al., 2021) and words in word embedding space (Mikolov et al., 2013).

The model in this study presents an algorithm of transforming locations in space into cognitive states in cognitive map, which embeds structured knowledge about spaces in general. Therefore, our model can be generalized to other sensory and cognitive domains that need to be charted and navigated with structured knowledge (Mikolov et al., 2013; Nau et al., 2018; Bao et al., 2019; Huang et al., 2021; Sosa and Giocomo, 2021; McNamee et al., 2021). Underlying this generality may reflect a hypothetic process of neuronal recycling (Dehaene and Cohen, 2007) that the development of grid cells to map a natural space through evolution may be gradually extended to chart spaces in other domains (Moser et al., 2017). In this way, the mapping of a variety of mental spaces with the same hexagonal metric system may help constructing a unified percept of the world efficiently (VanRullen and Kanai, 2021).

Materials and Methods

Participants

Two groups of right-handed university students with normal or corrected-to-normal vision were recruited. The first group, served as the experimental group, was 10 students from Peking University, and the second group contained ten students from Tsinghua University for replication purpose. Participants' age and sex were balanced between groups (experimental group: mean age = 22.10 [SD = 2.23], 5 females, 5 males; replication group: mean age = 23.9 [SD = 3.51], 5 females, 5 males). All participants had no history of psychiatric or neurological disorders and gave their written informed consent prior to the experiment, which was approved by the Research Ethics Committee of both universities.

Experimental design

Artificial object space. The greebles (Gauthier and Tarr, 1997) were created by 3ds Max (Autodesk, version 2022, http://www.autodesk.com) and the MAXScript codes from an open-source repository (TarrLab stimulus datasets, http://www.tarrlab.org/). The identity of individual greeble was determined by the lengths of its two features, named as "Loogit" and "Vacso" (Fig.1a, top). The names were arbitrary and had no relation to the object matching task (see below). We created 2025 greebles in total, each of which was equipped with a unique combination of Loogit's and Vacso's lengths. In this way, a 45-by-45 greeble space was constructed, with y-axis denoting Loogit's length and x-axis for Vacso's length (Fig. 1b). Therefore, in this 2-D artificial object space, each greeble represented a location. The greeble in the center of the space had equal length of Loogit and Vacso (i.e., prototype), which served as the target of the task. The ratio of a greeble parts' length over that of the prototype ranged from 0.12 to 1.88 with a step size of 0.04, which offered a smooth morphing experience when the participants adjusted parts' lengths.

Experimental procedure. In a trial of the object matching task, a greeble variant that was pseudo-randomly selected from greeble variants in the ring near the boundary of the space (red points in Fig.1b) was presented in the center of the screen after a 2.0-s fixation screen. The participants were instructed to adjust the lengths of the variant to match those of the prototype as quickly as possible by pressing buttons. There were four buttons, two of which were used to either stretching or shrinking one greeble part. When the participants felt that the adjusted variant completely matched the prototype, they either pressed the return key to end the trial (self-pace in behavioral training) or waited the time limit to run out (fixed time limit in fMRI scan) (see below). To ensure paths from variants to the prototype covered most of locations in the space, 24

variants were randomly selected with its angular precision of 1° in each 30° bin of the ring, which made up 288 variants evenly distributed in the ring. The accuracy of a trial was indexed by the distance between prototype's location and the location where the participant stopped. To encourage the participants to perform the task as accurately as possible, participants were informed that the amount of their monetary compensation increased exponentially with their performance. The task was programmed by Python (version 3.9) and Pygame (a library designed for writing video games, version 2.0, https://www.pygame.org/).

The task was conducted twice in two consecutive days, one for behavioral training and the other for fMRI scan. In day 1, the participants adjusted the length of variant's parts to match the prototype in a self-paced fashion. There were 8 sessions with 36 trials each, which made up 288 trials in total and took about 1 hour. The first 7 sessions provided feedback at the end of trials (e.g., "Loogit: 5 percent longer than the target") and the 8th session had no feedback for measuring behavioral performance. Before the training, the participants learned the prototype by watching it on the screen, and then performed at least 10 trials with feedback as practice. In day 2, the procedure was identical to that of day 1, except that (1) the self-pace was replaced with a 10-s time limit for each trial with a timer presented on top of screen (Fig.1a bottom), (2) no feedback was provided for all sessions, and (3) the participants finished the experiment in fMRI scanner. After the fMRI scan, the participants were debriefed, and none was aware of the existence of the object space or the analogy of the adjustment of parts' length as movement in space.

fMRI data acquisition

Imaging data were collected using a 3T Siemens Prisma scanner equipped with a 64- and 20-channel receiver head coil for the experimental group and replication group at PKU and Tsinghua Imaging Center, respectively. Functional data were acquired with a Multi-band Echo Planer imaging (EPI) sequence (TR = 2000 ms, TE = 30 ms, matrix size: 112 × 112 x 62, flip angle: 90°, resolution: 2 × 2 × 2.3 mm^3, number of slices: 62, slice thickness: 2 mm, gap = 0.3 mm, slice orientation: transversal). A high-resolution T1-weighted three-dimensional anatomical data set was collected for registration (MPRAGE, TR = 2530 ms, TE = 2.98 ms, matrix size: 448 x 512 x 192, flip angle: 7°, resolution: 0.5 × 0.5 × 1 mm3, number of slices: 192, slice thickness: 1 mm, slice orientation: sagittal). In the experimental group, stimuli were presented through a Sinorad LCD projector (Shenzhen Sinorad Medical Electronics) onto a 33-inch rear-projection screen. In the replication group, the stimuli were presented using a visual/audio stimulation system (Shenzhen Sinorad SA-9939) with a customized-design (Shenzhen Sinorad Medical Electronics Co., Ltd) MRI compatible 40" LED liquid crystal display. The resolution of screens was 1024 x 768, and the participants viewed the stimuli with an angled mirror mounted on the head coil.

fMRI data preprocessing

The BOLD signal time series of each scan session was preprocessed independently using FSL FEAT toolbox (FMRIB's Software Library, version 6.0, https://fsl.fmrib.ox.ac.uk/fsl/fslwiki; Woolrich et al. 2001; Woolrich et al. 2004). For each session, the time series were corrected for motion artefacts, slice time acquisition differences, geometrical distortion using the fieldmaps, and were high-pass filtered at 100 s. Spatial smoothing was further applied using a 5 mm FWHM Gaussian kernel. For group-level statistics, each session's functional data were registered to

participant's high-resolution anatomical image and were then normalized to a standard image (MNI152) using FSL FLIRT (Jenkinson and Smith, 2001). This procedure also resampled the functional voxels into a 2 x 2 x 2 mm resolution.

Analysis on hexagonal pattern in the ERC

To test whether the hexagonal pattern was present during the object matching task, we first computed the hexagonal orientation in the ERC for each participant using one half of the data (odd-numbered sessions). Second, we looked for the hexagonal pattern that modulated by participants' spatial orientation ranging from 1° to 360° in the object space after controlling for the effect of the hexagonal orientation (Doeller et al., 2010; Constantinescu et al., 2016).

Specifically, a GLM was created for each session to model the time series convolved with FSL Double-Gamma hemodynamic response function in the ERC with two parametric modulators: sin(6θ) and cos(6θ), where θ represents the spatial-orientation in the object space. The factor 6 means the 6-folded neural activity, which is rotational symmetric and aligned with spatial-orientation. The factor 5 and 7 were used as controls, which misaligned with the hexagonal pattern. Those estimated parameters β_{sine} and β_{cosine} were subsequently used to compute the hexagonal orientation φ, ranged from 0° to 59°, in each voxel and each participant, where $\varphi = [\arctan(\frac{\beta_{sine}}{\beta_{cosine}})]/6$. The hexagonal orientation Φ for each participant was derived by averaging the φ map across the odd-numbered sessions and voxels in the ERC.

Once the hexagonal pattern of neural activity in the ERC was identified, a second GLM was fitted to model the BOLD signals from the other half of the data (even-numbered sessions),

which was specified by a cosine parametric modulator: $\cos(6[\theta - \Phi])$. The activity map of the hexagonal pattern was derived by averaging the parametric map of this cosine modulator across the even-numbered sessions, which was further converted to z-score map before submitting to second-level analysis (Schlichting et al., 2015).

Analysis on spatial rhythm in the HPC

To examine whether there were spatial rhythms in BOLD signals in the Medial Temporal Lobe (MTL) that encompasses the ERC and HPC, we created a GLM to extract the activity map corresponding to each spatial orientation in the object space, and then performed voxel-based spectral analyses to identify robust rhythmic activity. Specifically, the spatial orientations, which were represented by paths from variants to the prototype, were down-sampled into bins with 15° each (e.g., the spatial orientations within -7.5° to 7.5°was label as 0° and the spatial orientations within 337.5° to 352.5°was label as 345°). This procedure resulted in a vector of 24 resampled spatial orientations (i.e., number of 15° bins in a ring), which could be used to identify spatial rhythms ranging from 0hz to 11hz. The timing of those spatial orientations was subsequently specified as parametric modulators in a GLM to model the BOLD signals. The generated activity maps for spatial orientations were then sorted in an ascending order (i.e., 0° to 345°) for each participant.

The time series of the spatial-orientation-dependent activity were extracted from each voxel of the MTL, and were then preprocessed by detrending, applying Hanning tapper, and zero-padding before submitting to spectral analyses performed by the Fast Discrete Fourier Transform (FFT) in the stats package of R (version 4.0, https://www.r-project.org/). (Landau and Fries, 2012).

After converting the time series into spatial frequency domain by FFT, we computed the absolute values of complex FFT outcome to derive the magnitude. This procedure further generated the 3D magnitude map for each frequency. The 95th percentile of pseudo t-statistic distribution (mean divided by SE across participants) in the MTL was computed as initial threshold to derive uncorrected clusters before second-level analyses. In addition, to examine the phase coupling between cortical regions, phases were also exacted as the ratio of the imagery part to the real part of FFT outcome using the inverse tangent function (atan2).

Anatomical masks

We manually delineated the bilateral ERC for ROI analyses on the MNI T1 brain with 2 mm resolution packaged by FSL using established protocols (Insausti et al. 1998), as well as a delineating software ITK-SNAP (Version 3.8, www.itksnap.org). The masks of HPC subfields were derived by an automated segmentation tool packaged by Freesurfer (Version 7.1, https://surfer.nmr.mgh.harvard.edu/; Iglesias et al., 2015), which segments the CA1 and CA2/3 area from hippocampal formation and further divides the mask into head and body parts. All masks were resampled to 2 mm resolution and aligned in standard MNI space.

Field size of place cells

To construct matrix P of HPC place cells used for the computational model and ANNs, parameters that determined the geometry of place field was selected based on the field size reported by a physiological study (Muller et al., 1987), where the field size was defined as the number of pixels inside the enclosure of place field Γ relative to the number of pixels over space χ. To do so, we extracted the field size distribution from 31 rats' place cells reported in Muller et al (1987) using

a data extraction tool (WebPlotDigitizer, version 4.5, https://automeris.io/WebPlotDigitizer/, Marin et al., 2017). This procedure suggested that the field size of place cells is approximately 2% relative to the space size (median value = 2.14%, and the values at 5th & 95th percentile are 0.98% & 4.87%, respectively). Besides this empirical field size, we also tested other field sizes (1%, 2%, 3% and 4%) in ANN simulation (Fig. S5).

Parameters for artificial neural network

One-layer CNN, RNN and MLP networks with 100 hidden units were trained using Torch (Version 1.9) and GeForce RTX 2070 SUPER graphics card on a PC. During training. Ten thousand images with the size of 100 × 100 pixels were used as input, each of which contains one unique point (i.e., one pixel) that represents a location in space consisting of 100 × 100 locations. The output was matrix *P* that represents cognitive map in the HPC, whose size was 1,000 (place cell) × 10,000 (locations in space). The value of matrix *P* was the Hadamard product of place cell's response field and 3Hz spatial rhythm in response to locations in space as shown in equation (1). Besides 3Hz, we also tested spatial rhythms ranging from 0 – 9 hz in ANN simulation.

The ANNs were optimized iteratively using stochastic gradient-descent algorithms (i.e., Adam) with a learning rate of 1e-4 and weight decay of 1e-5 for a total of 5,000 iterations (Loshchilov & Hutter, 2017; Walawalkar et al., 2020). After the training, the hexagonal pattern was examined by projecting units' activity map over the space. The gridness of the hexagonal pattern was calculated based on its autocorrelation map confined by the circular area around the outermost peak of six peaks closest to the map center, and the difference between the lowest correlation at 60° and 120°

and the highest correlation at 30°, 90°, and 150° was calculated to index gridness (Sargolini et al., 2006).

Statistical analysis

Hexagonal and spatial rhythmic signals were examined by cluster-based permutation using FSL randomise package (version 2.9, http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/Randomise), where an initial threshold of $p < 0.01$ was used to derive the uncorrected clusters. If no clusters were revealed, a liberal threshold of $p < 0.05$ was used. The reliability of uncorrected clusters was further tested using a non-parametric statistical inference that does not make assumptions about the distribution of the data (Nichols and Holmes 2002), where 1,000 random sign-flips were performed on the beta or FFT magnitude brain images. Clusters with size higher than 95% of the maximal suprathreshold clusters from permutation distribution were reported. Note that all the t-tests used in the present study were two-tailed.

The significance of gridness score was calculated for each hexagonal pattern generated by ANNs using field-shuffling-based permutation (Krupic et al., 2012) (Fig.S6). Specifically, we disrupted long-range spatial structure (spatial relationship among firing fields) by shuffling the firing fields of a given hexagonal pattern 100 times with the local structure preserved (the activity pattern of single field) to generate permutation distribution. The value in the 95th percentile of maximum gridscores was defined as the uncorrected threshold for the hexagonal pattern. Multiple comparison correction was applied to the uncorrected threshold by setting the maximum among spatial rhythms and place field sizes tested.

Data availability

The data that support the findings of this study are available from the lead corresponding author (J.L.) upon request.

Code availability

All analyzes reported in this work were conducted with customized codes written in Python (version 3.9), Pygame (version 2.0), Matlab (version 2019b), R (version 4.0), and BNU bash (version 3.2.57), those codes will be available from the corresponding author (J.L.) upon request. The neuroimaging and deep neural network toolbox used in this work are publicly available as following: FreeSurfer (v7.1, https://surfer.nmr.mgh.harvard.edu/), FSL (version 6.0, https://fsl.fmrib.ox.ac.uk/fsl/fslwiki;), and Torch (version 1.9, https://pytorch.org/).

Funding

This work was supported by the National Natural Science Foundation of China (31861143039), Tsinghua University Guoqiang Institute (2020GQG1016), Tsinghua University Qiyuan Laboratory, and Beijing Academy of Artificial Intelligence (BAAI).

Acknowledgments

We gratefully acknowledge Mr. Yue Wu and Ms. Yuannan Li for the discussion on computational modelling and simulation, Dr. Michael Tarr for sharing the code for greeble stimulus generation. Computational work was supported by High-performance Computing Platform of Peking University.

Conflict of interest

None declared

Author Contributions

J.L. conceived the study and provided advice. B.Z. conducted the fMRI study. Y.Z. built the computational model. B.Z. performed the neural network simulation. B.Z. and J.L. wrote the paper.

Reference

Amari, S. I. (1977). Dynamics of pattern formation in lateral-inhibition type neural fields. Biological cybernetics, 27(2), 77-87.

Burgess, N., Barry, C., & O'keefe, J. (2007). An oscillatory interference model of grid cell firing. Hippocampus, 17(9), 801-812.

Burak, Y., & Fiete, I. R. (2009). Accurate path integration in continuous attractor network models of grid cells. PLoS computational biology, 5(2), e1000291.

Bush, D., Barry, C., & Burgess, N. (2014). What do grid cells contribute to place cell firing?. Trends in neurosciences, 37(3), 136-145.

Bieri, K. W., Bobbitt, K. N., & Colgin, L. L. (2014). Slow and fast gamma rhythms coordinate different spatial coding modes in hippocampal place cells. Neuron, 82(3), 670-681.

Bush, D., Barry, C., Manson, D., & Burgess, N. (2015). Using grid cells for navigation. Neuron, 87(3), 507-520.

Bellmund, J. L., Gärdenfors, P., Moser, E. I., & Doeller, C. F. (2018). Navigating cognition: Spatial codes for human thinking. Science, 362(6415).

Behrens, T. E., Muller, T. H., Whittington, J. C., Mark, S., Baram, A. B., Stachenfeld, K. L., & Kurth-Nelson, Z. (2018). What is a cognitive map? Organizing knowledge for flexible behavior. Neuron, 100(2), 490-509.

Bao, X., Gjorgieva, E., Shanahan, L. K., Howard, J. D., Kahnt, T., & Gottfried, J. A. (2019). Grid-like neural representations support olfactory navigation of a two-dimensional odor space. Neuron, 102(5), 1066-1075.

Couey, J. J., Witoelar, A., Zhang, S. J., Zheng, K., Ye, J., Dunn, B., ... & Witter, M. P. (2013). Recurrent inhibitory circuitry as a mechanism for grid formation. Nature neuroscience, 16(3), 318-324.

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

De Valois, R. L., Albrecht, D. G., & Thorell, L. G. (1982). Spatial frequency selectivity of cells in macaque visual cortex. Vision research, 22(5), 545-559.

Dehaene, S., & Cohen, L. (2007). Cultural recycling of cortical maps. Neuron, 56(2), 384-398.

Doeller, C. F., Barry, C., & Burgess, N. (2010). Evidence for grid cells in a human memory network. Nature, 463(7281), 657-661.

Domnisoru, C., Kinkhabwala, A. A., & Tank, D. W. (2013). Membrane potential dynamics of grid cells. Nature, 495(7440), 199-204.

Dordek, Y., Soudry, D., Meir, R., & Derdikman, D. (2016). Extracting grid cell characteristics from place cell inputs using non-negative principal component analysis. Elife, 5, e10094.

Evans, T., Bicanski, A., Bush, D., & Burgess, N. (2016). How environment and self-motion combine in neural representations of space. The Journal of physiology, 594(22), 6535-6546.

Fyhn, M., Hafting, T., Witter, M. P., Moser, E. I., & Moser, M. B. (2008). Grid cells in mice. Hippocampus, 18(12), 1230-1238.

Fiebelkorn, I. C., Pinsk, M. A., & Kastner, S. (2018). A dynamic interplay within the frontoparietal network underlies rhythmic spatial attention. Neuron, 99(4), 842-853.

Gauthier, I., & Tarr, M. J. (1997). Becoming a “Greeble” expert: Exploring mechanisms for face recognition. Vision research, 37(12), 1673-1682.

Grill-Spector, K. (2003). The neural basis of object perception. Current opinion in neurobiology, 13(2), 159-166.

Guanella, A., Kiper, D., & Verschure, P. (2007). A model of grid cells based on a twisted torus topology. International journal of neural systems, 17(04), 231-240.

Hubel, D. H., & Wiesel, T. N. (1959). Receptive fields of single neurones in the cat's striate cortex. The Journal of physiology, 148(3), 574-591.

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. The Journal of physiology, 160(1), 106-154.

Hafting, T., Fyhn, M., Molden, S., Moser, M. B., & Moser, E. I. (2005). Microstructure of a spatial map in the entorhinal cortex. Nature, 436(7052), 801-806.

Huang, T., Zhen, Z., & Liu, J. (2021). Semantic relatedness emerges in deep convolutional neural networks designed for object recognition. Frontiers in computational neuroscience, 15, 16.

Insausti, R., Juottonen, K., Soininen, H., Insausti, A. M., Partanen, K., Vainio, P., ... & Pitkänen, A. (1998). MR volumetric analysis of the human entorhinal, perirhinal, and temporopolar cortices. American journal of neuroradiology, 19(4), 659-671.

Iglesias, J. E., Augustinack, J. C., Nguyen, K., Player, C. M., Player, A., Wright, M., ... & Alzheimer's Disease Neuroimaging Initiative. (2015). A computational atlas of the hippocampal formation using ex vivo, ultra-high resolution MRI: application to adaptive segmentation of in vivo MRI. Neuroimage, 115, 117-137.

Jenkinson, M., & Smith, S. (2001). A global optimisation method for robust affine registration of brain images. Medical image analysis, 5(2), 143-156.

Kant, I. (2006). The Cambridge companion to Kant and modern philosophy. Cambridge University Press.

Krupic, J., Burgess, N., & O'Keefe, J. (2012). Neural representations of location composed of spatially periodic bands. Science, 337(6096), 853-857.

Killian, N. J., Jutras, M. J., & Buffalo, E. A. (2012). A map of visual space in the primate entorhinal cortex. Nature, 491(7426), 761-764.

Landau, A. N., & Fries, P. (2012). Attention samples stimuli rhythmically. Current biology, 22(11), 1000-1004.

Loshchilov, I., & Hutter, F. (2017). Decoupled weight decay regularization. arXiv preprint arXiv:1711.05101.

Marr, D. (1982). Vision: A computational investigation into the human representation and processing of visual information. San Francisco: W.H. Freeman.

Muller, R. U., Kubie, J. L., & Ranck, J. B. (1987). Spatial firing patterns of hippocampal complex-spike cells in a fixed environment. Journal of Neuroscience, 7(7), 1935-1950.

McNaughton, B. L., Battaglia, F. P., Jensen, O., Moser, E. I., & Moser, M. B. (2006). Path integration and the neural basis of the'cognitive map'. Nature Reviews Neuroscience, 7(8), 663-678.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781.

Moser, E. I., Moser, M. B., & McNaughton, B. L. (2017). Spatial representation in the hippocampal formation: a history. Nature neuroscience, 20(11), 1448-1464.

Marin, F., Rohatgi, A., & Charlot, S. (2017). WebPlotDigitizer, a polyvalent and free software to extract spectra from old astronomical publications: application to ultraviolet spectropolarimetry. arXiv preprint arXiv:1708.02025.

McNamee, D. C., Stachenfeld, K. L., Botvinick, M. M., & Gershman, S. J. (2021). Flexible modulation of sequence generation in the entorhinal–hippocampal system. Nature Neuroscience, 24(6), 851-862.

Nichols, T. E., & Holmes, A. P. (2002). Nonparametric permutation tests for functional neuroimaging: a primer with examples. Human brain mapping, 15(1), 1-25.

Nau, M., Schröder, T. N., Bellmund, J. L., & Doeller, C. F. (2018). Hexadirectional coding of visual space in human entorhinal cortex. Nature neuroscience, 21(2), 188-190.

O'Keefe, J., & Dostrovsky, J. (1971). The hippocampus as a spatial map: preliminary evidence from unit activity in the freely-moving rat. Brain research.

Rich, P. D., Liaw, H. P., & Lee, A. K. (2014). Large environments reveal the statistical structure governing hippocampal representations. Science, 345(6198), 814-817.

Samsonovich, A., & McNaughton, B. L. (1997). Path integration and cognitive mapping in a continuous attractor neural network model. Journal of Neuroscience, 17(15), 5900-5920.

Solstad, T., Moser, E. I., & Einevoll, G. T. (2006). From grid cells to place cells: a mathematical model. Hippocampus, 16(12), 1026-1031.

Sargolini, F., Fyhn, M., Hafting, T., McNaughton, B. L., Witter, M. P., Moser, M. B., & Moser, E. I. (2006). Conjunctive representation of position, direction, and velocity in entorhinal cortex. Science, 312(5774), 758-762.

Stemmler, M., Mathis, A., & Herz, A. V. (2015). Connecting multiple spatial scales to decode the population activity of grid cells. Science Advances, 1(11), e1500816.

Schlichting, M. L., Mumford, J. A., & Preston, A. R. (2015). Learning-related representational changes reveal dissociable integration and separation signatures in the hippocampus and prefrontal cortex. Nature communications, 6(1), 1-10.

Sorscher, B., Mel, G., Ganguli, S., & Ocko, S. A. (2019). A unified theory for the origin of grid cells through the lens of pattern formation.

Sorscher, B., Mel, G. C., Ocko, S. A., Giocomo, L., & Ganguli, S. (2020). A unified theory for the computational and mechanistic origins of grid cells. bioRxiv.

Sosa, M., & Giocomo, L. M. (2021). Navigating for reward. Nature Reviews Neuroscience, 1-16.

Tolman, E. C. (1948). Cognitive maps in rats and men. Psychological review, 55(4), 189.

Von der Malsburg, C. (1999). The what and why of binding: the modeler's perspective. Neuron, 24(1), 95-104.

van Groen, T., Miettinen, P., & Kadish, I. (2003). The entorhinal cortex of the mouse: organization of the projection to the hippocampal formation. Hippocampus, 13(1), 133-149.

VanRullen, R., & Kanai, R. (2021). Deep learning and the Global Workspace Theory. Trends in Neurosciences.

Witter, M. P., & Amaral, D. G. (1991). Entorhinal cortex of the monkey: V. Projections to the dentate gyrus, hippocampus, and subicular complex. Journal of Comparative Neurology, 307(3), 437-459.

Woolrich, M. W., Ripley, B. D., Brady, M., & Smith, S. M. (2001). Temporal Autocorrelation in Univariate Linear Modeling of FMRI Data. NeuroImage, 14(6), 1370–1386. http://doi.org/10.1006/nimg.2001.0931

Woolrich, M. W., Behrens, T. E. J., & Smith, S. M. (2004). Constrained linear basis sets for HRF modelling using Variational Bayes. NeuroImage, 21(4), 1748–1761. http://doi.org/10.1016/j.neuroimage.2003.12.024

Walawalkar, D., Shen, Z., Liu, Z., & Savvides, M. (2020). Attentive cutmix: An enhanced data augmentation approach for deep learning based image classification. arXiv preprint arXiv:2003.13048.

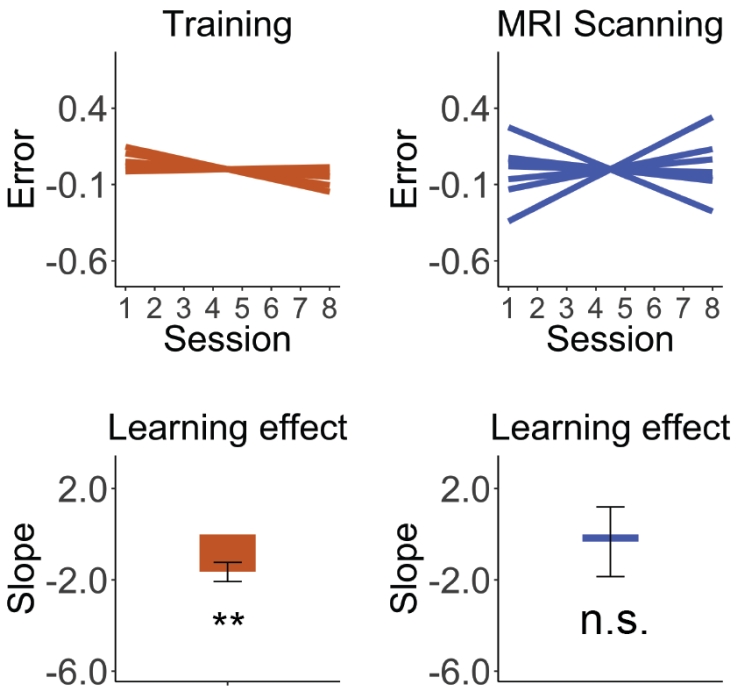
Yartsev, M. M. (2013). Space bats: multidimensional spatial representation in the bat. Science, 342(6158), 573-574.

Supplemental Material

Origin of hexagonal metric of grid cells in navigating spaces

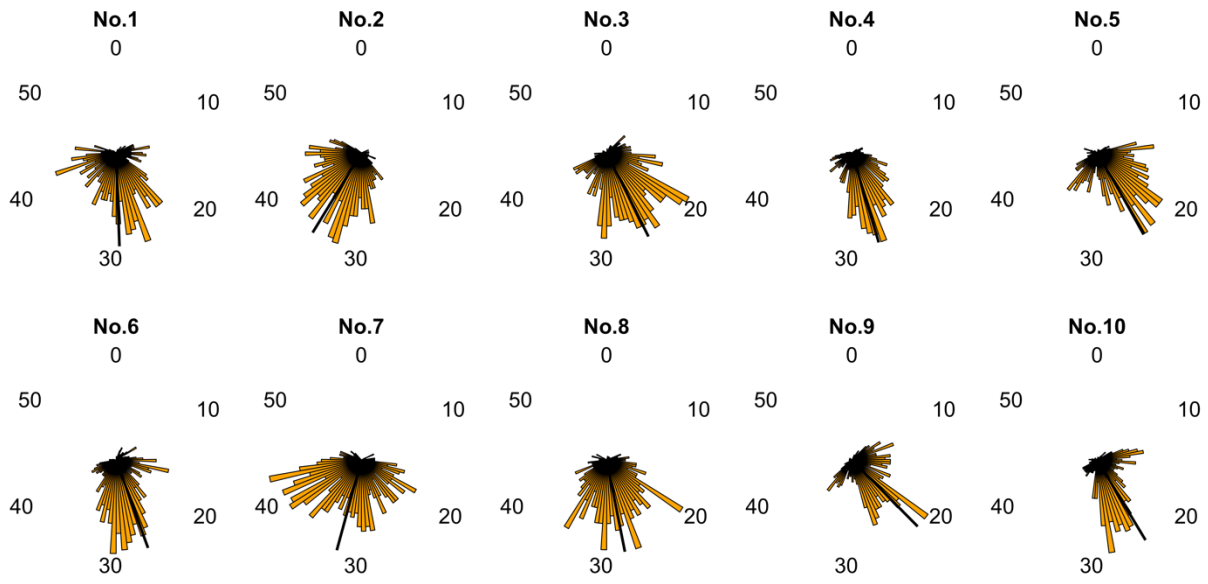
Bo Zhang, Yiyuan Zhang and Jia Liu

Figure S1. Behavioral results.



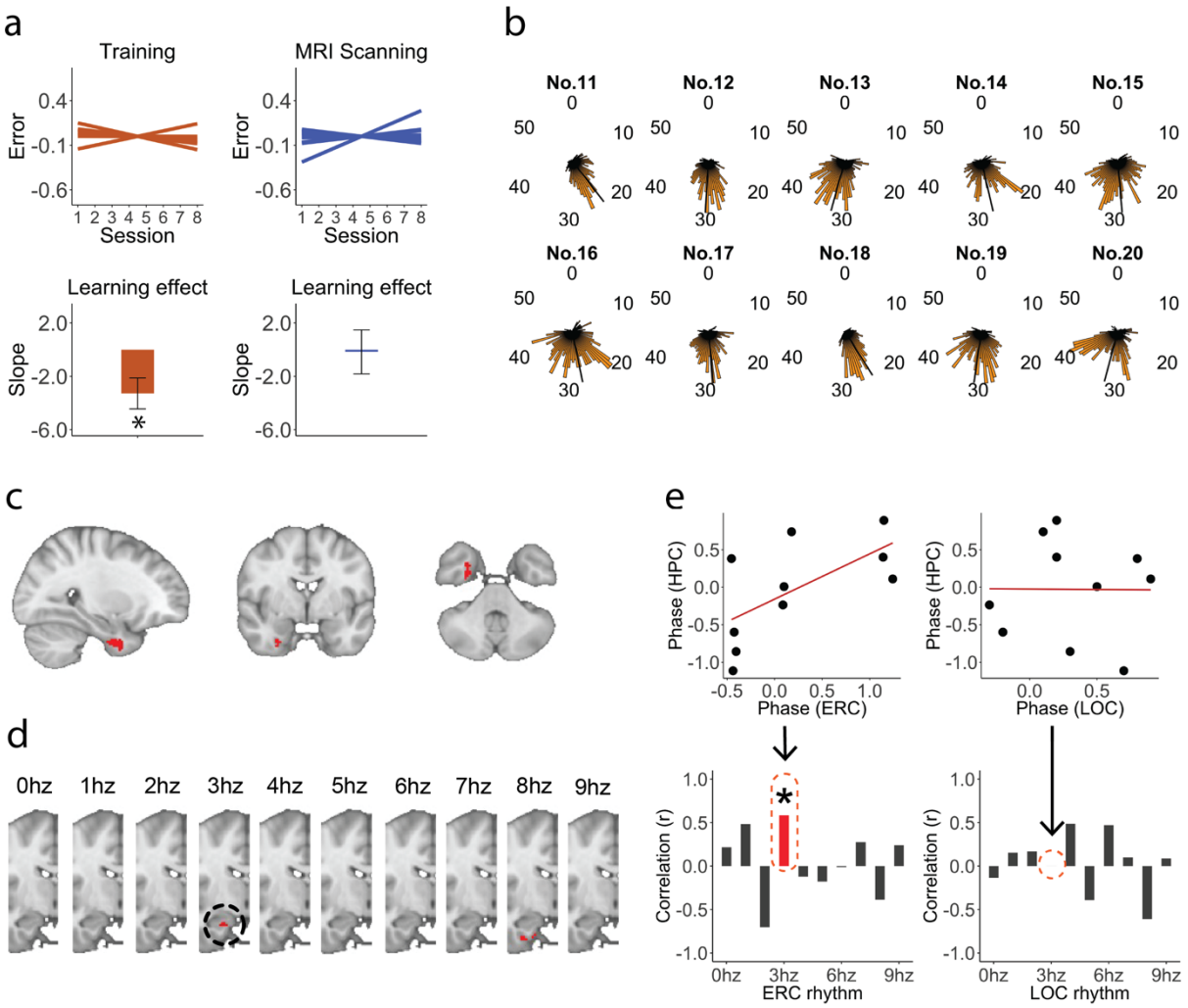
In day 1, participants were getting familiar with the task to minimize the learning effect during MRI scan in day 2. Errors in object matching were defined as the distance between prototype's location and the location where the participants stopped. A significant negative slope was revealed in errors across sessions ($t(9) = -3.9$, $p = 0.003$) in day 1, showing a significant learning effect. No learning effect was observed during fMRI scan ($t(9) < 1$) in day 2, ruling out the possibility of learning effect on the formation of the hexagonal pattern in the ERC and the 3Hz spatial rhythm in the HPC. **: $p < 0.01$; n.s.: not significant. Error bars indicate SEM.

Figure S2. Distribution of hexagonal orientations in the ERC



The uniformity of hexagonal orientations, which was computed from each voxel in the ERC, was tested by Rayleigh test. All participants showed significantly clustered hexagonal orientations ($p < 0.05$, Bonferroni corrected). Each plot is for one participant, where the length of yellow bar indicates the number of voxels and the black line denotes the mean of hexagonal orientations.

Figure S3. Behavioral and MRI results of replication group



(a) A significant increase in behavioral performance of object matching was found in day 1 ($t(9) = -2.83$, $p = 0.01$), while no significant change in performance was observed during fMRI scan in day 2 ($t(9) < 1$). (b) The hexagonal orientations, tested by Rayleigh test, showed significant uniformity for all participants ($p < 0.05$, Bonferroni corrected). (c) A significant cluster was revealed in the right ERC for 6-fold hexagonal modulation (initial threshold: $p < 0.01$; SVC for multiple comparisons: $p < 0.05$), but not for 5- and 7-fold signal modulation. (d) A spectral analysis revealed a significant cluster in the right HPC showing significant spatial rhythm of 3hz (initial threshold on pseudo t-value: $p < 0.05$; SVC for multiple comparison: $p < 0.05$; Peak MNI coordinates: 22, -10, -24). (e) Significant phase coupling between the HPC and the ERC was found only at the spatial rhythm of 3hz (Pearson r = 0.64, $p = 0.04$). No significant correlation was found between the HPC and the LOC (r = 0.00).

Figure S4. Hexagonal patterns of all hidden units that had significant gridness score for each ANN.

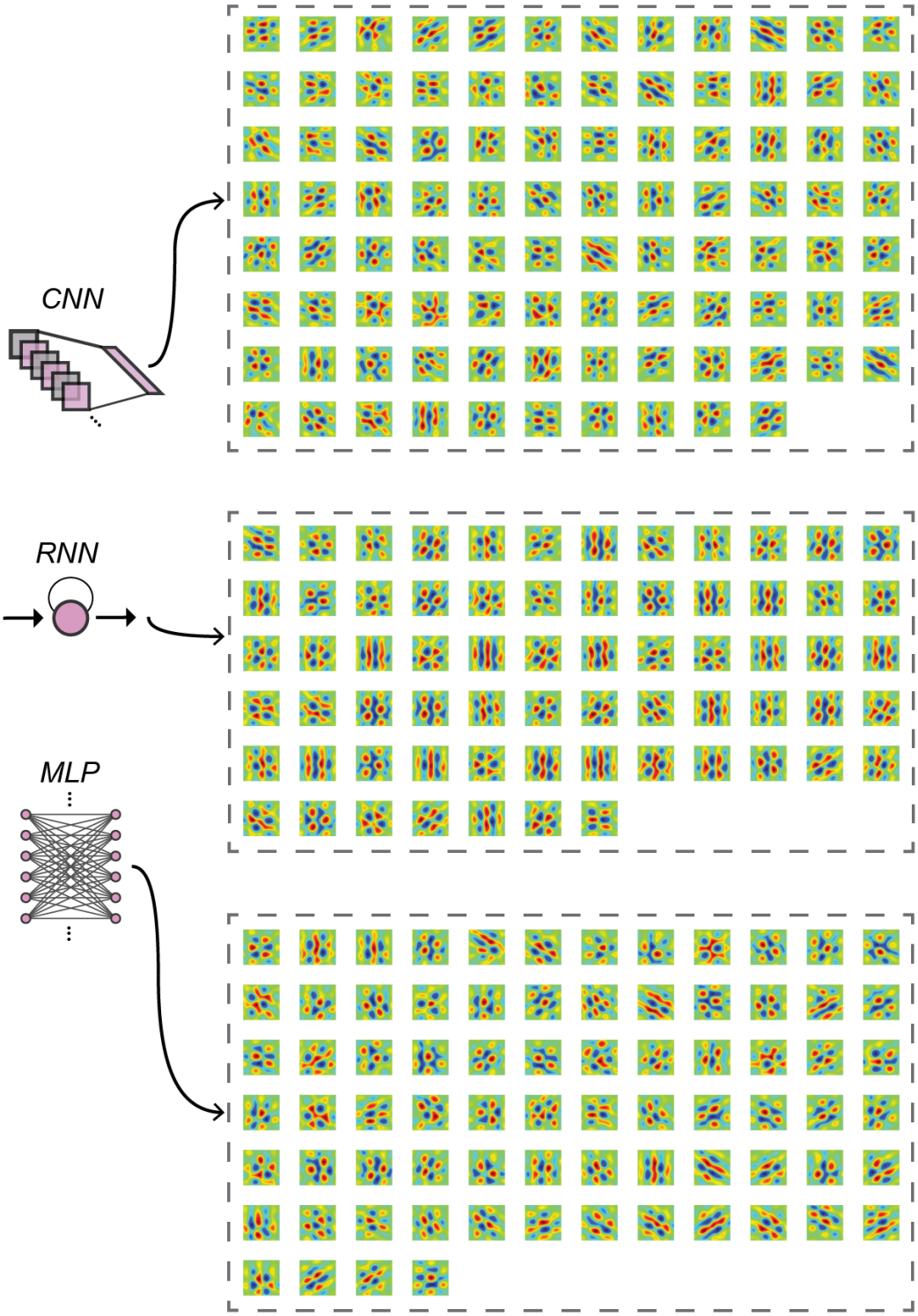
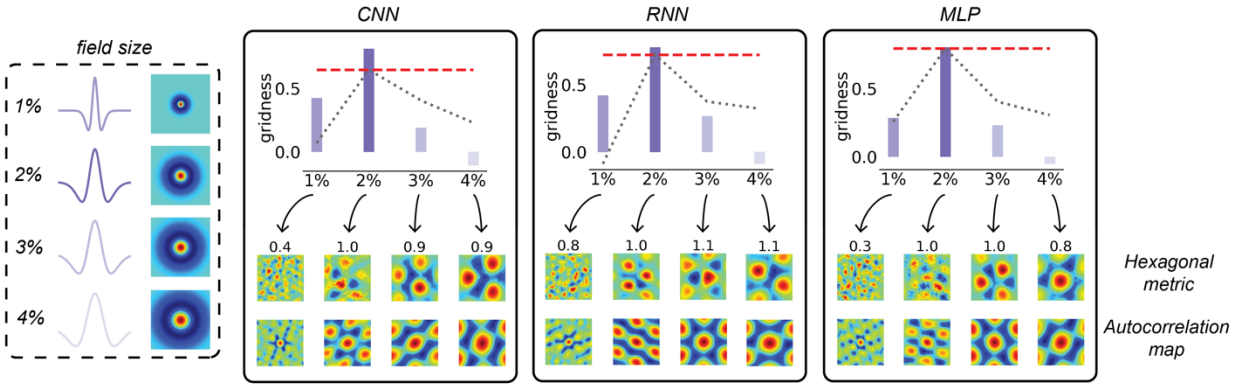
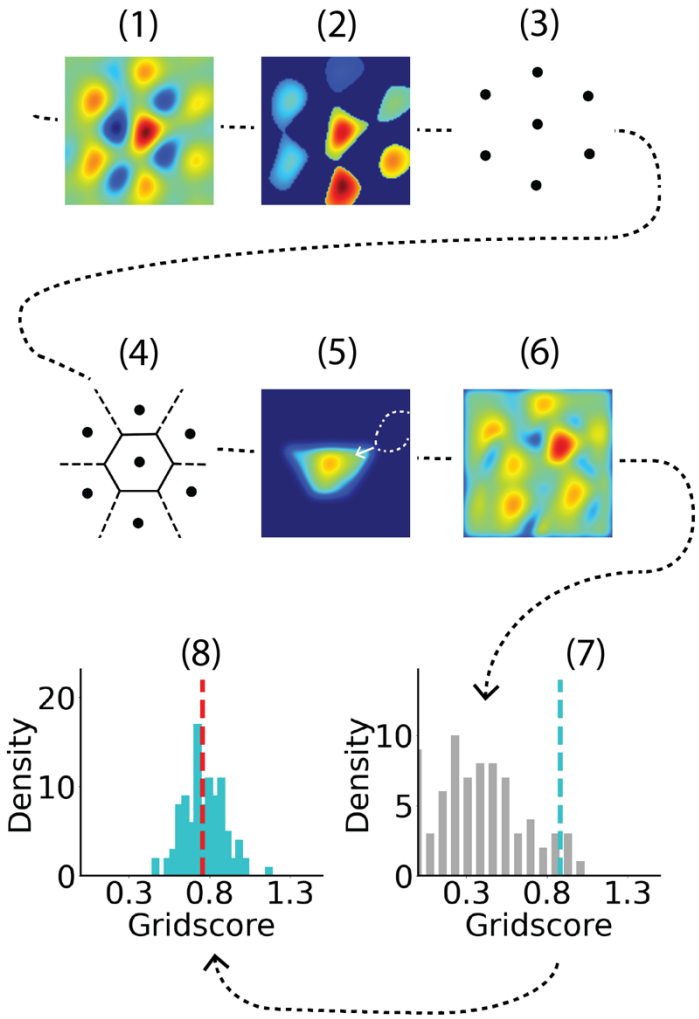


Figure S5. The size of place field on hexagonal pattern



In neural network simulation, we also tested field sizes of place cells other than 2% in generating the hexagonal pattern of grid cells. Left, the field sizes of 1%, 2%, 3%, and 4% relative to the size of space were tested with the parameters of standard deviation σ_1 = [4, 8, 10, 12] and σ_2 = [8, 16, 20, 24] and amplitude α = [1] and β = [0.6], respectively. Right, as shown in upper panel, significant hexagonal pattern, measured by the mean gridscore of 100 units in the hidden layer of ANNs that exceeded the threshold from the field shuffling procedure, was only observed for the field size of 2% ($p < 0.05$, permutation test, corrected for multiple comparison). Gray and red dash lines indicate uncorrected and corrected thresholds of gridness, respectively. However, carefully examination of each hidden unit in ANNs with field sizes of 1%, 3%, and 4% revealed that the activity of a small number of units showed the hexagonal pattern. First row: exemplar gridscore among 100 hidden units and their activity pattern; second row: the autocorrelation map. This simulation result suggests that the field size of place cells in the HPC was important for generating a large number of units with regular hexagonal pattern in the ERC.

Figure S6. Procedure for shuffling hexagonal fields



(1) An exemplar of hexagonal pattern, (2) the local structure by applying a threshold of median value, (3) the peaks of local structure, (4) the segmentation of local structure by Voronoi polygons, (5) the field that was randomly rotated and positioned during shuffling, (6) an exemplar of shuffled hexagonal pattern, (7) the distribution of shuffled gridscores for a hexagonal pattern, the dash line represents the threshold at the 95th percentile of the maximum gridscore, (8) the distribution of the thresholds from all hexagonal patterns with the mean value represented by red dash line.