Manuscript submitted to eLife

A model of a panoramic visual representation in the dorsal visual pathway: the case of spatial reorientation and memory-based search

Tianyi Li[1], **Angelo Arleo**[1], **Denis Sheynikhovich**[1*]

***For correspondence:**
denis.sheynikhovich@upmc.fr (DS)

[1]Sorbonne Université, INSERM, CNRS, Institut de la Vision, 17 rue Moreau, F-75012 Paris, France.

Abstract Primates are primarily visual animals and understanding how visual information is processed on its way to memory structures is crucial to the understanding of how memory-based visuospatial behavior is generated. Recent imaging data demonstrate the existence of scene-sensitive areas in the dorsal visual path that are likely to combine visual information from successive egocentric views, while behavioral evidence indicates the memory of surrounding visual space in extraretinal coordinates. The present work focuses on the computational nature of a panoramic representation that is proposed to link visual and mnemonic functions during natural behavior. In a spiking artificial neuron network model of the dorsal visual path it is shown how time-integration of spatial views can give rise to such a representation and how it can subsequently be used to perform memory-based spatial reorientation and visual search. More generally, the model predicts a common role of view-based allocentric memory storage in spatial and not-spatial mnemonic behaviors.

Introduction

Recent breathtaking advances in our understanding of rodent hippocampal memory system pave the way for elucidating the organization of human spatial memory (***Burgess, 2014***; ***Moser et al., 2017***). One major difference between primates and rodents is the role of vision for behavior. Primates are much more visual animals than rodents and understanding the link between primate visual and medial temporal lobe (MTL) memory structures is an important and largely unexplored open question (***Meister and Buffalo, 2016***). Experimental evidence indicates the existence of functional and anatomical connections between these structures. Functional connections are demonstrated by two principal lines of studies. First, visual behavior is informed by memory as demonstrated by studies of novelty preference in both monkeys and humans (***Wilson and Goldman-Rakic, 1994***; ***Manns et al., 2000***; ***Jutras and Buffalo, 2010a***). In the novelty preference paradigm, the memory is assessed from looking time: well memorized stimuli are looked at less than novel ones. The specific role of MTL structures in this phenomenon is derived from results showing a decreased novelty preference after MTL lesions or in patients suffering from mild cognitive impairment or Alzheimer's disease, often associated with MTL dysfunction (***McKee and Squire, 1993***; ***Crutcher et al., 2009***; ***Zola et al., 2013***). In monkeys, restricted lesions of hippocampal and/or

parahippocampal cortices also decreased novelty preference (***Zola et al., 2000***; ***Pascalis et al., 2009***; ***Bachevalier et al., 2015***). Second, the link between visual and MTL structures is manifested in coherent neural activities in the two structures. For example, activity of single MTL neurons is modulated by visual saccades (***Sobotka et al., 1997***), the onset of visual stimuli strongly affects hippocampal neural responses (***Jutras and Buffalo, 2010a***) and hippocampal theta oscillations are reset by eye movements (***Jutras and Buffalo, 2010b***; ***Hoffman et al., 2013***).

Anatomical connections between visual and memory structures have recently been characterized in the novel framework of the occipital–parietal–MTL pathway of visuospatial processing (***Kravitz et al., 2011***). There are three principal stages of information processing in this pathway (***Figure 1***A). First, the occipito-parietal circuit processes visual information through visual areas V1-V6 an egocentric (retinal) frame of reference. Successive information processing in these areas is thought to extract visual features of increasing complexity, including motion and depth cues and relay this information to the parietal cortex. Second, a complex network of interconnected parietal structures relays highly-processed visual cues to support executive, motor and spatial-navigation functions. These structures include the medial, ventral and lateral intraparietal areas (MIP, VIP, LIP) strongly linked with eye movements processing; the middle temporal and medial superior temporal (MT, MST) thought to extract high-level visual motion cues; and the caudal part of the inferior parietal lobule (cIPL), the main relay stage on the way to the medial temporal lobe. The cIPL sends direct projections to the CA1 of the hippocampus as well as to the nearby parahippocampal cortex (PHC). In addition, it sends indirect projections to the same structures via the posterior cingulate cortex (PCC) and the retrosplenial cortex (RSC). Within this complex network, neurons at different neurobiological sites have been reported to code space in a world- or object-centred reference frames (***Duhamel et al., 1997***; ***Snyder et al., 1998***; ***Chafee et al., 2007***). Moreover, both PCC and RSC have been repeatedly linked to coordinate transformation between egocentric and allocentric frames of reference (***Vogt et al., 1992***; ***Burgess, 2008***; ***Epstein and Vass, 2014***). Importantly, information processing in this pathway is strongly affected by directional information thought to be provided by a network of head-direction cells residing in several brain areas, including RSC (***Taube, 2007***). Finally, medial temporal lobe, and in particular the hippocampus, play a key role in constructing an allocentric representation of space in primates (***Hori et al., 2003***; ***Ekstrom et al., 2003***).

Given functional and anatomical connections between visual and memory structures, the question arises as to the nature of neuronal representations in the dorsal visual path. In addition to the well-established role of parieto-retrosplenial networks in coordinate transformations (***Andersen et al., 1993***; ***Snyder et al., 1998***; ***Salinas and Abbott, 2001***; ***Pouget et al., 2002***; ***Byrne et al., 2007***), a largely unexplored question concerns the existence of an extra-retinal neural map of the remembered visual space (***Hayhoe et al., 2003***; ***Tatler and Land, 2011***; ***Land, 2014***). That the task-related visual retinotopic space is remembered has been suggested by studies showing that when asking to recall a recent visual content, eye movements (on a blank screen) closely reflected spatial relations of remembered images (***Brandt and Stark, 1997***; ***Johansson and Johansson, 2014***). Moreover, preventing subjects from making eye movements decreased recall performance (***Johansson and Johansson, 2014***; ***Laeng et al., 2014***). That not only the retinal egocentric space is remembered but also extra-retinal map of surrounding space is stored in memory is demonstrated in studies showing that during natural behavior human subjects direct saccades toward extra-retinal locations, suggesting that these locations are represented in memory, potentially in an allocentric frame of reference (***Land et al., 1999***; ***Hayhoe et al., 2003***; ***Golomb et al., 2011***; ***Melcher and Morrone, 2015***; ***Robertson et al., 2016***). Even though suggested by the above studies, the nature of such an extra-retinal map and neural mechanisms underlying its construction and storage are currently unknown.

The present modeling study addresses the question of how such an allocentric representation of surrounding visual space can be constructed and stored by the dorsal visual path – MTL networks. We propose that the existence of such a representation relies on short-term memory linking

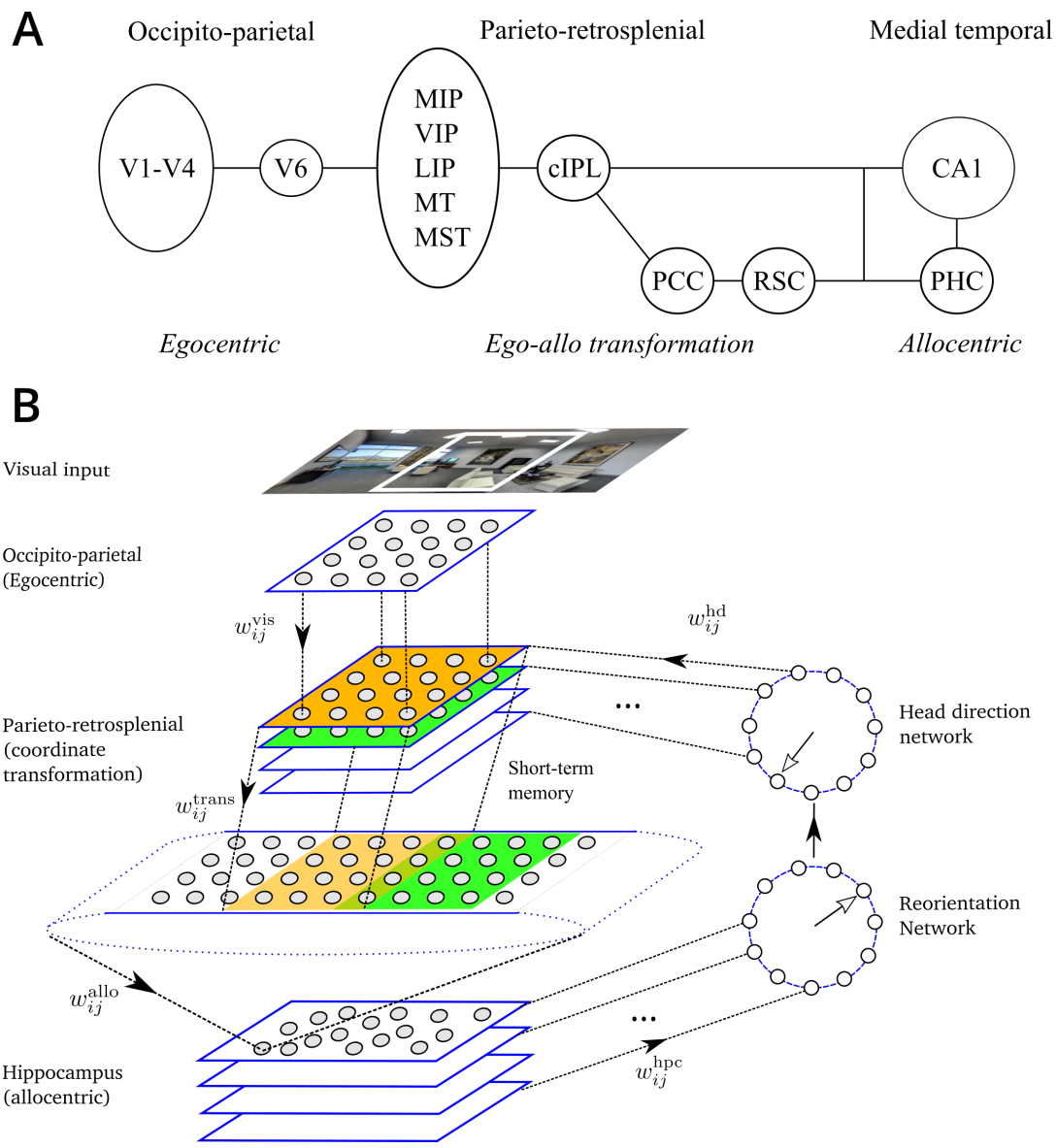

Figure 1. Model. A. Dorsal visual pathway of visuospatial information processing in primates (see text for details). B. Schematic representation of the model. Visual features present in the limited visual field constitute the model input. The model network is composed of 6 modules: (1) Occipito-parietal (egocentric); (2) Head-direction network; (3) Parieto-retrosplenial transformation network consists of the coordinate-transformation network and the output layer, which encodes visual features in an allocentric directional frame and spans 2π; (4) Hippocampus; (5) Reorientation network. Projections from the occipito-parietal (visual) areas to the transformation network are topographic. Each head-direction cell activates the corresponding layer of the transformation network. Projections from the different layers of the transformation network to the parieto-retrosplenial output layer are also organized according to head direction: any two layers project topographically to overlapping portions of the output population shifted according to head direction. Synapses between the transformation network and the parietal output network are endowed with short-term memory. Different hippocampal subpopulations project to different neurons in the reorientation network, which in turn corrects head direction signal. Full arrows represent the flow of information in the network. Open arrows represent direction signals in the head direction and reorientation networks.

successive egocentric views and we study how the long-term memory of allocentric visual space can affect behavior in spatial and non-spatial experimental paradigms. In particular, our results suggest that allocentric memory effects during spatial reorientation and memory-based visual guidance

tasks can be explained by the existence of such a network.

Methods

The model is a spiking neuron network constructed to reflect information processing steps thought to be performed by successive stages of neuronal processing in the primate dorsal visual path described above (***Figure 1***A). To reflect in a simplified way the main processing stages in the pathway, our model of the dorsal pathway is composed of 5 main modules or subnetworks (***Figure 1***B). First, the module representing information processing in the occipito-parietal circuit essentially applies a set of Gabor-like orientation filters to the incoming visual images, a standard assumption for basic V1 processing. We do not model eye movements, and assume that a retinotopic visual representations obtained at the level of V1 has been remapped, by the time it arrives into the parietal cortex, to a head-fixed representation by taking into account eye position information (***Duhamel et al., 1997***; ***Snyder et al., 1998***; ***Pouget et al., 2002***). Even though gaze independent, this head-fixed representation is egocentric, or view-dependent, in the sense it depends on the position and orientation the modeled animal (i.e., its head) in space. Second, we model the directional sense by a network of cells whose activity is approximately Gaussian around their preferred orientations (***Taube, 2007***) and that is sending projections to the parietal cortex (***Brotchie et al., 1995***; ***Snyder et al., 1998***). Third, both the activities of the egocentric network and the head direction signal converge onto the network modeling the role of the parieto-retrosplenial network in coordinate transformation. This transformation network uses head direction to convert egocentric visual representations into a head-orientation-independent, or world-fixed representation. This coordinate transformation is done essentially by the same mechanism as the retinotopic-to-head-fixed conversion mentioned above, but in contrast to previous models it does so using low-level topographic visual information. The resulting orientation-independent visual representation is often referred to as spatiotopic, or allocentric, since visual features are determined a world-fixed directional reference frame. Fourth, the allocentric output of the parieto-retrosplenial network arrives to the hippocampus, modeled by a network of cells that learn, by a competitive mechanism, allocentric visual patterns provided by the parietal network. As will be clear from the following, in the context of spatial navigation these cells can be considered as place cells, whereas in a non-spatial context they can be considered as representing memorised visual stimuli. Finally, the reorientation module associates allocentric memories with directional reference frame and feeds back to the head direction cells. The activity of this network represents the correction signal for self-orientation. When the memorized information corresponds to the newly arrived one, the correction signal is zero, whereas in the case of disorientation or in response to specific manipulations of visual cues, it can provide fast adjustment of the self-orientation signal. In the Results section we show that a similar reorientation mechanism can be responsible for behavioral decisions in spatial, as well as non-spatial tasks in primates.

Occipito-parietal input circuit

The occipito-parietal network is modeled by a single rectangular sheet of $N_x \times N_y$ visual neurons, uniformly covering the visual field. In all simulations, except Simulation 6 below, the size of the visual field was limited to $160 \times 100°$, approximately representing that of a primate. The activities of these visual neurons are computed in four steps. First, input images are convolved (using OpenCV `filter2D()` function) with Gabor filters of 4 different orientations $(0, 90°, 180°, 270°)$ at 2 spatial frequencies (0.5 cpd, 2.5 cpd), chosen so as to detect visual features in simulated experiments. Second, the 8 convolution images are discretized with $N_x \times N_y$ grid, and the maximal response at each position is chosen, producing an array of $N_x N_y$ filter responses. These operations are assumed to roughly mimic retinotopic V1 processing (***Heeger, 1992***), transformed into a head-fixed reference frame using eye-position information. Third, the vector of filter activities at time t is normalized to have maximal value of unity. Fourth, a population of $N_{vis} = N_x N_y$ Poisson neurons is created with mean rates given by the activity of the corresponding filters scaled by the constant maximal rate

A_{vis} (see ***Table 1*** for the values of all parameters in the model). For a Poisson neuron with rate r, the probability of emitting a spike during a small period of time δt is equal to $r\delta t$ (***Gerstner et al., 2014***).

Head direction

The head direction network is composed of $N_{\text{hd}} = 36$ Poisson neurons organized in a circle, such that neurons' preferred directions ϕ_k are uniformly distributed between 0 and 2π. The tuning curves of the modeled head-direction neurons are Gaussian with maximum rate A_{hd} and width $\sigma_{\text{hd}} = 8°$. Thus, the rate of head-direction neuron k when the model animal's head is oriented in the direction ϕ is given by

$$r_k^{\text{hd}} = A_{\text{hd}} \exp\left(-\frac{(\phi - \phi_k)^2}{\sigma_{\text{hd}}^2}\right) \tag{1}$$

Such a network generates a Gaussian activity profile centered around ϕ. Our model does not explicitly implement a line attractor dynamics hypothesized to support head direction signal (***Zhang, 1996***), but it is consistent with it. Head direction cells have been found in several brain areas in rodents and primates (see ***Taube, 2007***, for review), and there is evidence that parietal cortex receives head direction signals (***Brotchie et al., 1995***).

Parietal transformation network

The parietal transformation network is inspired by previous models (***Becker and Burgess, 2001***; ***Byrne et al., 2007***) but in contrast to them it operates directly on activities of the Gabor-like visual cells. The transformation of coordinates between the head-fixed and world-fixed coordinates is performed by multiple subpopulations of leaky integrate-and-fire (LIF) neurons organized as two-dimensional layers of neurons (see ***Figure 1***). Neurons in each layer of the transformation network are in a one-to-one relationship with the visual population and so at each moment t each transformation layer receives a copy of the egocentric (head-fixed) visual input. Therefore, the number of neurons in each transformation layer is equal to N_{vis}. Apart from the visual input, the transformation network also receives input from the population of head direction cells. There is a topographic relationship between the sub-populations of the transformation network and different head directions: each head-direction cell sends excitatory projections to neurons only in one associated layer of the transformation network. Thus, input from head-direction cells strongly activates only a small subset of transformation layers which transmit visual information to the downstream population. More specifically, only the layers which are associated with head directions close to the actual orientation of the head are active. The number of layers in the transformation network is then equal to N_{hd}, giving the total number of neurons in the transformation network $N_{\text{trans}} = N_{\text{vis}} N_{\text{hd}}$.

Thus, in a k-th layer of the transformation network, the membrane potential $v_i(t)$ of the LIF neuron i in is governed by the following equation (omitting the layer index for clarity):

$$\tau_{\text{m}} \frac{dv_i}{dt} = V_{\text{rest}} - v_i + g_i^{\text{ex}}(t)(E_{\text{ex}} - v_i) + g_i^{\text{in}}(t)(E_{\text{in}} - v_i) + R_{\text{m}} I_{\text{ext}} \tag{2}$$

with the membrane time constant τ_m, resting potential V_{rest}, excitatory and inhibitory reversal potentials E_{ex} and E_{in}, as well as the membrane resistance R_{m}. When the membrane potential reaches threshold V_{th}, the neuron fires an action potential. At the same time, v_i is reset to V_{reset} and the neuron enters the absolute refractory period Δ_{abs} during which it cannot emit spikes. A constant external current I_{ext} is added to each neuron to simulate baseline activity induced by other (unspecified) neurons from the network.

The excitatory conductance in these neurons depends only on the visual input (and thus is independent from k). It is modeled as a combination of α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid (AMPA) and N-methyl-d-aspartate (NMDA) receptor activation $g_i^{\text{ex}} = (1-\alpha) g_i^{\text{ampa}} + \alpha g_i^{\text{nmda}}$, that are

described by

$$\tau_{\text{ampa}} \frac{dg_i^{\text{ampa}}}{dt} = -g_i^{\text{ampa}} + \tau_{\text{ampa}} \sum_{j\in\{vis\}} w_{ij}^{\text{vis}} s_j(t) \quad (3)$$

$$\tau_{\text{nmda}} \frac{dg_i^{\text{nmda}}}{dt} = -g_i^{\text{nmda}} + g_i^{\text{ampa}} \quad (4)$$

where the index j runs over input (visual) neurons connected to it, w_{ij}^{vis} are the connection weights and $s_j(t) = 1$ if a presynaptic spike arrives at time t and $s_j(t) = 0$ otherwise. Constants τ_{ampa} and τ_{nmda} determine the time scales of receptor activation.

In contrast, the inhibitory conductance depends only on the head-direction cells and ensures that a small subset of transformation layers (i.e. those associated with nearby head directions) are active. To implement it, we employ a simple scheme in which all transformation layer neurons are self-inhibitory, and this inhibition is counteracted by the excitatory input from the head-direction cells. Thus, the inhibitory conductance of the i-th neuron in the k-th layer is given by

$$\tau_{\text{gaba}} \frac{dg_i^{\text{in}}}{dt} = -g_i^{\text{in}} + G_{\text{inh}} + \tau_{\text{gaba}} \sum_{k\in\{hd\}} w_{ik}^{\text{hd}} s_k(t) \quad (5)$$

where G_{inh} is the constant maximum amount of self-inhibition and w_{ik}^{hd} are the synaptic weights of connections from the head-direction cells. In the current implementation, there is one-to-one correspondence between the head-direction cells and the layers of the transformation network, so $w_{ik} = 1$ only for associated head-direction cell ϕ_k and $w_{ik} = 0$ otherwise.

All layers of the transformation network project to the parietal output population, which codes image features in an allocentric (world-fixed) directional frame. The parietal output population is represented by a two-dimensional neuronal sheet spanning $360 \times 100°$, that is a full panoramic view. It is encoded by a grid of $N_{\text{x}}^{\text{allo}} \times N_{\text{y}}^{\text{allo}}$ neurons. Each layer of the transformation network projects to a portion of the population according to the head direction associated with it associated with this layer (see ***Figure 1***). Since any two nearby layers of the transformation network are associated with head directions shifted relative to each other by $360°/N_{\text{hd}} = 10°$, the overlap between their projections on the parietal output layer is $140°$.

Thus, at each moment in time, a spiking representation of the current visual stream (i.e. a spiking copy of the visual input, gated by the head direction cells) arrives to the allocentric neurons spatially shifted according to the current head direction. For example, if two egocentric views (each spanning $160°$) are observed at head directions $-45°$ and $45°$ with respect to an arbitrary north direction, these two views arrive at the allocentric population spatially shifted relative to one another by $90°$, so that the activated neurons in the allocentric population span $230°$. To ensure that subsequent snapshots are accumulated in time (e.g. during head rotation), the synapses between neurons in the transformation layers and the allocentric population are endowed with short-term memory, implemented by a prolonged activation of NMDA receptors (***Durstewitz et al., 2000***). Such synapses result in a sustained activity of allocentric output neurons during a period of time sufficient for downstream plasticity mechanism to store information from accumulated snapshots.

The membrane potential of the i-th neuron in the allocentric output population is governed by ***Equation 2*** with the synaptic conductance terms determined as follows. First, the excitatory AMPA conductance is given by ***Equation 3*** but with the input provided by transformation network neurons via weights w_{ij}^{trans}. Second, the NMDA conductance is described by ***Equation 4***, but with the synaptic time scale increased by a factor of 6. This is done to ensure sustained activation of the output neurons upon changes in the visual input. Third, inhibitory input is set to zero for these neurons.

Learning the weights in the transformation network

The connection weights w_{ij}^{vis} from the visual neurons to the parietal transformation cells and w_{ij}^{trans} from the parietal transformation cells to the parietal output neurons are assumed to be learned during development by a supervised mechanism, similar to the one proposed to occur during

sensory-motor transformation learning (***Zipser and Andersen, 1988***; ***Salinas and Abbott, 1995***). In this models it is proposed that when an object is seen (i.e. its retinal position and an associated gaze direction are given), grasping the object by hand (that operates w.r.t. the body-fixed reference frame) provides a teaching signal to learn the coordinate transformation. A similar process is assumed to occur here, but instead of learning body-based coordinates using gaze direction, the model learns world-fixed coordinates using head direction.

More specifically, synaptic weights in the coordinate-transformation network were set by the following procedure. First, the network was presented with an edge-like stimulus at a random orientation and at a randomly chosen location in the visual field. Second, upon the stimulus presentation, the head direction was fixed at a randomly chosen angle ϕ. Third, neurons in the transformation layers associated with the chosen head direction were activated with the average firing rates equal to the rates of the corresponding visual neurons, while neurons in the parietal output layer were activated with the same average rates but shifted according to the chosen head direction (representing the teaching signal). Fourth, the synaptic weights in the network were set according to the Hebbian prescription:

$$w_{ij}^{\text{vis}} = r_i^{\text{trans}} r_j^{\text{vis}} \tag{6}$$

$$w_{ij}^{\text{trans}} = r_i^{\text{trans}} r_j^{\text{allo}} \tag{7}$$

where r_i^{vis}, r_i^{trans} and r_i^{allo} are the mean firing rates of the corresponding visual neurons, transformation network neurons and parietal output neurons, respectively. Fifth, the weight vector of each neuron was normalized to have the unity norm. This procedure has been performed for edge-like stimuli at 4 different orientations (corresponding to 4 Gabor filter orientations), placed in the locations spanning the whole visual field and at head directions spanning 360°. Synaptic weights (***Equation 6-7***) were fixed to the learned values prior to all the simulation presented here. No updates were performed on these weights during the simulations.

Hippocampal neurons

As a result of the upstream processing, neuronal input to the hippocampus represents visual features in an allocentric directional frame. Neurons in the parietal output population are connected in an all-to-all fashion to the population of modeled hippocampal cells and the connection weights that are updated during learning according to an spike-timing-dependent plasticity (STDP) rule below. In addition, lateral inhibition between hippocampal neurons ensures a soft winner-take-all dynamics, such that sufficiently different patterns in the visual input become associated with small distinct subpopulations of hippocampal neurons.

Thus, the membrane equation of the i-th hippocampal neurons is given by ***Equation 2***. The excitatory conductances are given by ***Equation 3-4***, but with the input provided by the parietal output neurons via weights w_{ij}^{allo}. Upon the initial entry to a novel environment these weights are initialized to small random values. During learning, the amount of synaptic modification induced by a single pair of pre- and post-synaptic spikes is given by

$$\frac{dw_{ij}^{\text{allo}}}{dt} = G_{\max}\left[a_j^{\text{pre}} s_i(t) - a_i^{\text{post}} s_j(t)\right] \tag{8}$$

where $s_i(t)$ and $s_j(t)$ detect pre- and post-synaptic spikes, respectively, and

$$\begin{aligned} \frac{da_j^{\text{pre}}}{dt} &= -\frac{a_j^{\text{pre}}}{\tau_{\text{pre}}} + A_+ s_j(t) \\ \frac{da_i^{\text{post}}}{dt} &= -\frac{a_i^{\text{post}}}{\tau_{\text{post}}} + A_- s_i(t) \end{aligned} \tag{9}$$

The inhibitory conductance of the hippocampal neuron is governed by the following equation:

$$\tau_{\text{gaba}} \frac{dg_i^{\text{in}}}{dt} = -g_i^{\text{in}} + \tau_{\text{gaba}} \sum_{j\in\{\text{hpc}\}} w_{ij}^{\text{inh}} s_j(t) \tag{10}$$

in which τ_{gaba} determines the time scale of synaptic inhibition as before, and the weights $w_{ij}^{\text{inh}} = W_{\text{inh}}$ are constant and ensure that each hippocampal neuron inhibits all other hippocampal neurons proportionally to its activity.

The hippocampal circuit is complex and consists of several interconnected populations. In our simple model of hippocampal activity we consider only the first stage of hippocampal processing of visual information that is likely to be the CA1, which receives direct projections from the entorhinal cortex, an input gateway to the hippocampus.

Reorientation network

During one continuous experimental trial (e.g. an exploration trial in novel environment or an observation of a novel image on the screen), the reference frame for head direction is fixed and all processing operations in the network are performed with respect to the origin of this reference frame. In particular, an allocentric information stored by the hippocampus as a result of the trial can be correctly used for future action only if the origin of the reference frame is stored with it. Therefore, if in a subsequent trial, the actions to be performed require memory of the previous one, the network should be able to recover the original directional reference (this of course can happen only the visual information received at the start of the trial is considered familiar). Reorientation is the process by which the origin of the stored reference frame is recovered.

Our model of this process rests on the assumption that it is automatic, fast, bottom-up, and does not require costly object/landmark processing. The support for this assumption comes from a large body of reorientation studies in many animal species including primates, showing that object identities are ignored during reorientation (***Cheng and Newcombe, 2005***). The conditions in which most of the reorientation studies were performed usually are such that there is no single conspicuous point-like cue in the environment that can be reliable associated with a reference direction. For example, in many studies the directional cues come from the geometric layout of the experimental room. Lesion studies in rats suggest that reorientation in such conditions requires an intact hippocampus (***McGregor et al., 2004***). Furthermore, we propose that this reorientation network is active all the time, in contrast to being consciously "turned on" when the animal "feels disoriented". Therefore, we expect that its effects can be observed even when no specific disorientation procedure was performed. In particular, we suggest in the Results that a manipulation of objects on the screen can result in automatic corrections of directional sense that can be observed during visual search.

The reorientation network in the model is organized similarly to the head-direction network and consists of N_{re} neurons with preferred positions uniformly distributed on a circle. Therefore, the difference between two nearby reorientation cells is $\Delta\phi = 2\pi / N_{\text{re}}$. The membrane potential of the i-th reorientation neuron is described by the LIF equation (***Equation 2***). Excitatory conductances are described by ***Equation 3-4*** with the input to the neuron provided by hippocampal place cells via weights w_{ij}^{hpc}. There is no inhibition in the network, and so the inhibitory conductance is set to 0. The ability of the network to perform reorientation is determined by afferent connection weights from the hippocampal cells, which are determined as follows.

Since all allocentric information learned during a trial is linked to the same directional frame, all hippocampal cells learned during the trial are connected to a single neuron of the reorientation network, the one with the preferred direction $0°$ (***Figure 2***). The connection weights between the hippocampal cells and the neuron are updated using STDP rule, ***Equation 8-9*** (this is not essential for the model to work, so that setting the weights to a constant value will give similar results). Once the training trial is finished, N_{re} copies of the learned hippocampal population are created, each corresponding to a separate neuron in the reorientation network. In each copy, all cells have the same input and output weights as the corresponding cells in the original population, but their connection profile is different. In particular, the copy that corresponds to the reorientation neuron with preferred direction $\Delta\phi$ is connected to pre-synaptic cells are shifted by the same angle in the topographically-organized allocentric layer (***Figure 2***). In machine learning literature, this technique

is called "weight sharing" and it allows to achieve translation invariance for detection of objects in images. Here, we apply a similar technique in order to detect familiar snapshots and head direction associated with them.

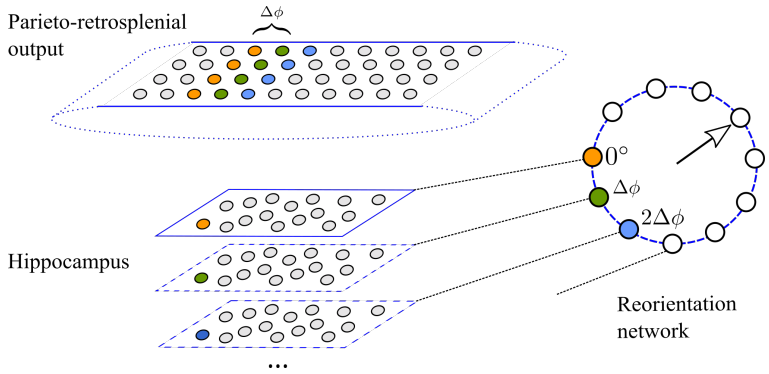

Figure 2. Implementation of the reorientation network. Top: the output population of the parieto-retrosplenial network. Bottom: hippocampal cells. The population outlined by full lines is the original population learned during training. As a result of learning, the hippocampal cell shown in orange is connected to the presynaptic cells of the same color (connection weights not shown). All cells in the original population are connected to a single cell ($o°$) in the reorientation network (Right). The hippocampal populations outlined by dashed lines are copies of the original population that implement weight sharing: the hippocampal cell shown in green (blue) has the same connection weights as the orange cell, but it is connected to pre- and post-synaptic cells shifted by $\Delta\phi$ ($2\Delta\phi$). The number of copies of the original hippocampal population is the same as the number of neurons in the reorientation network.

Suppose, for example, that as a result of learning during a trial, a hippocampal cell is associated with 4 presynaptic cells in the output layer of the transformation network (cells shown in orange in ***Figure 2***). Suppose further that during an inter-trial interval the head direction network has drifted (or was externally manipulated), so that at the start of the new trial the internal sense of direction is off by $2\Delta\phi$. When the animal sees the same visual pattern again, it will be projected onto the allocentric layer shifted by the same amount (blue cells in ***Figure 2***). This will in turn cause the hippocampal subpopulation that includes the blue cell to be most strongly active, such that the activity peak of the reorientation network signals the orientation error. The reorientation is then performed by readjusting the head direction network to minimize the reorientation error. In the current implementation this is done algorithmically by subtracting the error signal from the actual head direction, but it can also be implemented by attractor dynamics in the head direction layer.

Simulation details

The spiking artificial neural network model described above was implemented using Python 2.7 and Brian 2 spiking neural network simulator (***Stimberg et al., 2019***). The time step for neuronal simulation was set to 1 ms, while the sampling rate of visual information was 10 Hz, according to the proposals relating oscillatory brain rhythms in the range 6–10 Hz to information sampling (***Hasselmo et al., 2002***; ***Busch and VanRullen, 2010***). At the start of each simulation, the weights w_{ij}^{allo} and w_{ij}^{hpc} were initialized to small random values (the other weights were trained as described earlier and fixed for all simulations), see ***Figure 1***B. Parameters of the model are listed in ***Table 1***, and the sections below provide additional details of all simulations.

Simulation 1: Egocentric-allocentric transformation

The first simulation was inspired by the study of ***Snyder et al.*** (***1998***), in which monkeys observed visual stimuli at identical retinal locations, but for different orientations of the head with respect to the world, in order to assess whether parietal neurons were modulated by the allocentric head direction. Thus, in this simulation, the head direction angle ϕ was varied from $-50°$ to $50°$ in 100 sessions. For each trial of a session, the mean rates of the head-direction neurons were calculated

Parameter	Value	Description
	Neuron numbers	
$N_x \times N_y$	80×50	Parieto-occipital network size
N_{hd}	36	Head direction network size
$N_x^{allo} \times N_y^{allo}$	180×50	Parietal output layer size
N_{re}	36	Reorientation network size
	Mean amplitudes in the input populations	
A_{vis}	100	Spikes/s., Maximum rate of the parieto-occiptal network
A_{hd}	100	Spikes/s., Maximum rate of the head-direction network
	Parameters of the LIF model	
V_{rest}	-65	mV, Resting potential
V_{th}	-55	mV, Spiking threshold
V_{reset}	-65	mV, Reset potential
E_{ex}	0	mV, Excitatory reversal potential
E_{in}	-80	mV, Inhibitory reversal potential
E_{in}	250	mΩ, Membrane resistance
Δ_{abs}	1^{a-c}, 2^{d}	ms, Absolute refractory period
α	$0.9^{a,b}$, $0.3^{c,d}$	Balance between AMPA and NMDA receptor
τ_{ampa}	5	ms, AMPA receptor time scale
τ_{nmda}	$100^{a,c,d}$ 600^{b}	ms, NMDA receptor time scale
τ_x	2.5	ms, NMDA receptor time scale
τ_m	$10^{a,c,d}$, 20^{b}	ms, Membrane time scale
τ_{gaba}	10	ms, GABA receptor time scale
I_{ext}	20^{a-c}, 40^{d}	mA, External input current
G_{inh}	2	Self-inhibitory conductance
	STDP	
G_{max}	0.05^{c}, 0.1^{d}	Maximal weight change
A_+	0.005	Maximal potentiation amplitude
A_-	A_+ x 1.05	Maximal depression amplitude
τ_{pre}	20	ms, Potentiation time scale
τ_{post}	15^{c}, 17.5^{d}	ms, Depression time scale
	Other parameters	
σ_{hd}	8°	Tuning curve width of head direction cells
W_{inh}	1.0	Lateral inhibition weight in the hippocampal population

Table 1. Parameters of the model. a, Occipito-parietal circuit. b, Parieto-retrosplenial transformation network. c, Hippocampus. d, Reorientation network.

according to ***Equation 1*** and fixed for the rest of the trial. The stimulus (vertical black bar, width: 10°) was shifted horizontally across the midline of the visual field (160×100°) from left to right in 1° steps, such that it remained at each position for 100ms. The neuronal spikes were recorded from the occipito-parietal network, the parieto-retrosplenial transformation network and its output layer, for each stimulus position across 10 trials per session. Mean firing rates were then calculated from these data.

Simulation 2: Accumulation of successive views using short-term synaptic memory

The aim of the second simulation was to illustrate the synaptic mechanism for an integration of successive visual snapshots in time, instrumental for spatial coding. We model a monkey that remains in the same spatial location and turns its head from left to right. Thus, the model was presented with a set of 9 successive overlapping views (160×100°) taken from a panoramic (360×100°) image, 100ms per view. Initial head direction was arbitrarily set to 0°.

Simulation 3: Encoding of allocentric visual information during spatial exploration

In the third simulation we studied the role of temporal accumulation of visual information for spatial coding. The model ran through a square 3D environment (area: 10×10 m, wall height 6 m)for about 10 min so as to cover uniformly its area. The visual input was provided by a cylindrical camera ($160 \times 100°$) placed at the location of the model animal. At each spatial location 9 successive views of the environment were taken in different directions (as in the Simulation 2). The vector of mean firing rates of the occipito-parietal neurons at a single spatial location and orientation constituted the egocentric population vector. The mean firing rates of the the parieto-retrosplenial output neurons at each location constituted the allocentric population vector (this population vector is independent from orientation as a result of coordinate transformation). To compare spatial information content in the two populations, we first estimated intrinsic dimensionality of the two sets of population vectors. This was performed using two recent state-of-the art methods: DANCo (***Ceruti et al., 2014***), as implemented by the `intrinsicDimension` R package, and ID_fit (***Granata and Carnevale, 2016***). For both methods, the principal parameter affecting dimensionality estimation is the number of neighbors for each point in the set that is used to make local estimates of the manifold dimension. Second, we used two different methods to visualize the structure of the low-dimensional manifold: Isomap (***Tenenbaum et al., 2000***) and t-SNE (***van der Maaten and Hinton, 2008***). To extract principal axes of the manifold, we used PCA on the data points projected on two principal dimensions provided by the above methods. We chose the parameter values for which the visualized manifold best approximates the original space. We then determined a set of points (i.e. population vectors) that lie close to the principal axes of the manifold and visualized them in the original environment. If the manifold structure corresponds well to the spatial structure of the underlying environment, the principal axes of the manifold should lie close to the principal axes of the environment.

Simulation 4: Visual responses of hippocampal neurons in an image memorization task

This simulation was inspired by the study of ***Jutras and Buffalo*** (***2010a***) in which a large set of novel visual stimuli was presented to monkeys on a computer screen. Neuronal activity in the hippocampal formation in response to the visual stimuli was recorded. One of the results of this study suggested that hippocampal neurons encode stimulus novelty in their firing rates. To simulate this result, we presented to the model 100 novel stimuli randomly chosen from the dataset retrieved from http://www.vision.caltech.edu/Image_Datasets/Caltech101). The stimuli (resized to 160×100 pixels) were shown to the model successively in one continuous session (500ms stimulus presentation time + 1000ms inter-trial interval with no stimuli) and the activities of the hippocampal neurons during learning were recorded.

Simulation 5: Spatial reorientation

In this simulation of the experiment of ***Gouteux et al.*** (***2001***), the testing room was a rectangular 3D environment with area 20×10 m and wall height 6m. In the "No cues" task the only visual features in the room were provided by the outlines of the walls. In the other 3 tasks, a square visual cue was presented in the middle of one of the walls with the edge length equal to 1/6 (small cue), 1/3 (medium cue) or 1/2 (large cue) of the environment width. Each task consisted of two phases, exploration and reorientation. During the exploration phase the modeled animal uniformly explored the environment, as in Simulation 3. The reorientation phase composed multiple trials. At the beginning of each trial, the model was placed at one of spatial locations covering the environment in a uniform grid. At each of these locations, 9 successive views were taken. Reorientation performance was assessed in two ways: (i) only the first view at each location was used for reorientation; (ii) successive views accumulated over 60 successive positions were used for reorientation.

Simulation 6: Memory-based visual search

In this simulation we used a dataset of visual images used in the study by ***Fiehler et al.*** (***2014***). This dataset consists of 18 image sets corresponding to 18 different arrangements of the same 6 objects (mug, plate, egg, jam, butter, espresso cooker). Each set includes a control image (all objects on the table in their initial positions) and images in which one of the objects is missing (target object) and one or more other objects displaced to the left or to the right. In the simulation we used only a subset of all images in a set that included either 1, 3 or 5 of the objects mentioned above displaced either to the left or to the right (referred to as "local" condition in ***Fiehler et al., 2014***), giving rise to 6 experimental conditions. In each condition, there were 18 test images of displaced objects, plus the associated control images. Taking into account the distance between the animal and the screen as well as the size of the image (provided by ***Fiehler et al.*** (***2014***)), we calculated the size of the image in degrees of visual field. We then determined a rectangular portion of the image ($30 \times 15°$) that included all objects in initial and displaced positions in all images. The contents of this area served as an input to the model. Thus, in this simulation the spatial resolution of the visual input was higher than in the previous simulations as the visual field of the model was smaller, but the size of the input network was kept the same.

During each simulation trial, the image of objects in initial positions was first presented to the network during 2000 ms and stored by the hippocampal cells. The image of displaced objects (in one of the 6 conditions above) was subsequently presented to the network for the same amount of time and the orientation error was read out from the mean firing rates of the reorientation network.

Results

We first show that properties of neuronal firing along the simulated neural pathway from the visual cortex to the hippocampus reflect those of biological neurons along the pathway. We then demonstrate how backward projections from the hippocampus to the head direction network, can explain hippocampal influence on head direction during spatial reorientation and memory-based visual search.

Visual and parietal model neurons encode sensory representations in distinct reference frames

We start with a characterization of modeled dorsal-visual path neurons in the case when a simulated animal is assumed to sit in front of a screen and is free to rotate its head (***Duhamel et al., 1997***; ***Snyder et al., 1998***, for simplicity, we assume that rotation occurs only in the horizontal plane). The firing rate of occipito-parietal (input) neurons and the output parietal neurons as a function of the allocentric position of a visual stimulus (i.e. a vertical bar moving horizontally across the visual field) was measured for two different head directions (***Figure 3***A,B). For a neuron in the input population, a change in head direction induces the corresponding change of the receptive field of the neuron, since its receptive field shifts together with the head along the allocentric position axis (***Figure 3***C). In contrast, for a parietal output neuron, a change in head direction does not influence the position of its receptive field, which remains fixed in an allocentric frame (***Figure 3***D). To show that this is also true on the population level, we measured, for all visual input cells and all parietal output cells, the amount of shift in its receptive field position as a function of head direction shift, while the head was rotated from $-50°$ to $50°$. For cells in the occipito-parietal visual area, the average linear slope of the dependence is close to 1, whereas in the allocentric parietal population the average slope is close to 0 (***Figure 3***E), meaning that these two populations encode the visual stimulus in the two different reference frames: head-fixed and world-fixed. These properties of model neurons reproduce well-known monkey data showing that different sub-populations of parietal cortex neurons encode visual features in the two reference frames (***Duhamel et al., 1997***; ***Snyder et al., 1998***).

The receptive fields of the intermediate neurons of the coordinate transformation network

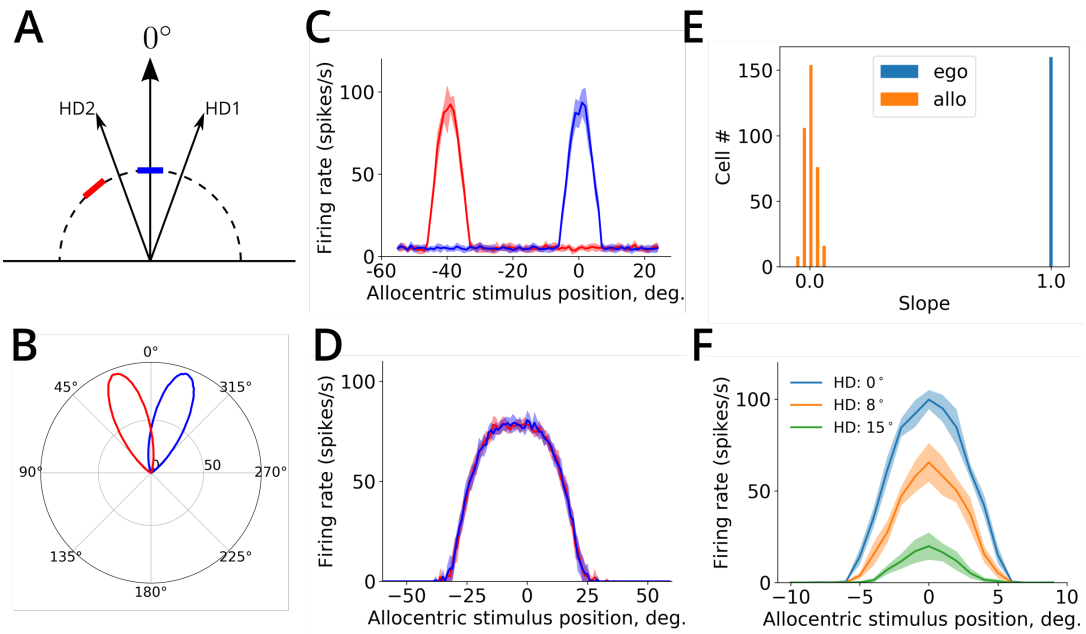

Figure 3. Properties of neurons in the coordinate-transformation network. A. A schematic representation of the receptive field of one input visual input neuron at two head directions (HD1 and HD2). The position of the receptive field of the neuron is shown by the blue and red bar for HD1 and HD2, respectively. B. The population activity of head direction cells in the model at 20° (HD1) and -20° (HD2). C. Tuning curves of an input visual neuron (±SD) for the two head directions represented in B. D. Tuning curves of an allocentric output neuron for the same head directions. E. Histograms show the distributions of the linear dependence slopes between the shift in the receptive field position and the shift in head direction, for egocentric (in blue) and allocentric (in orange) neuronal populations. F. Transformation network neurons are gain-modulated by head direction. Stimulus tuning curves of the same neuron for three different head directions are shown.

exhibit gain modulation by head direction (***Figure 3***F), as do monkey parietal neurons (***Snyder et al., 1998***).The hypothesis of reference-frame conversion via gain modulation has been extensively studied in both experimental and theoretical work, in the context of sensory-motor coordination during vision-guided reaching (***Avillac et al., 2005***; ***Pouget and Sejnowski, 1997***; ***Salinas and Abbott, 2001***). While coordinate-transformation processes involved in the two cases are conceptually similar, the underlying neuronal computations can differ substantially, because the former requires simultaneous remapping for the whole visual field, while the latter is limited to the computation of coordinates for a single target location (i.e. a representation of the point-like reaching target). This difference limits the use of noise-reducing attractor-like dynamics that is an essential component in point-based sensory-motor transformation models (***Pouget et al., 2002***), because in full-field transformation the information and noise are mixed together in a single visual input stream.

Spatial coding using temporal accumulation of successive views

Because of a limited view field, at each moment in time the simulated animal can directly observe only a restricted portion of visual environment (i.e. a visual snapshot, see ***Figure 4***A,B). That these snapshot-like representations are represented in memory, has been demonstrated in a number of studies showing viewpoint-dependent memory representations (***Diwadkar and McNamara, 1997***; ***Christou and Bülthoff, 1999***; ***Gaunet et al., 2001***). Moreover, experimental evidence suggests that visual information can be accumulated from successive snapshots during e.g. head rotation, giving rise to a panoramic-like representation of the surrounding environment that can inform future goal-oriented behavior (***Tatler et al., 2003***; ***Oliva et al., 2004***; ***Golomb et al., 2011***; ***Robertson et al., 2016***). A candidate neural mechanism for implementing such integration is short-term memory, i.e. the ability of a neuron to sustain stimulus-related activity for a short period of time (***Goldman-Rakic,***

1995). In our model, this is implemented by sustained firing via prolonged NMDA receptor activation (***Figure 4***C). Combined with STDP learning rule in the connections between the parietal output neurons and the hippocampus, this mechanism ensures that a time-integrated sequence of visual snapshots is stored in the synapses to hippocampal neurons. In particular, head rotation results in a temporarily activated panoramic representation in the population of output parietal neurons that project to CA1. STDP in these synapses ensures that these panoramic representations are stored in the synapses to downstream CA1 neurons (***Figure 4***D).

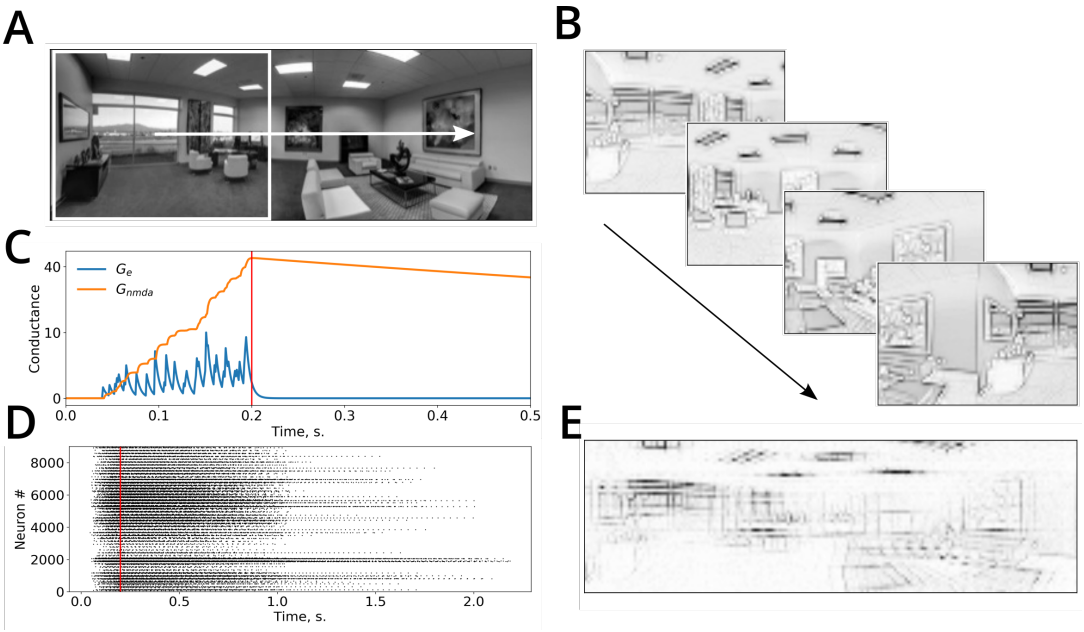

Figure 4. Temporal accumulation of successive visual snapshots in the model. A. A panoramic image of an environment superimposed with the visual field of the simulated animal (white rectangle). The white arrow shows the direction of visual scan path. B. Several successive visual snapshots along the scan path shown in A are represented by mean firing rates of the occipito-parietal (egocentric) network. C. An example of the evolution of AMPA and NMDA receptor conductances of parieto-retrosplenial output neurons as a function of time. Stimulus onset: $t = 0$, stimulus offset: $t = 200$ms (red line). D. Raster plot of spiking activities of the output neurons showing short-term memory in this network. An input is presented at time 0 and is switched off at the time shown by the red vertical line. The neurons remain active after stimulus offset due NMDA-receptor mediated short-term memory. E. Synaptic weight matrix of a single hippocampal neuron after learning stores the activity of the parieto-retrosplenial output layer accumulated over several successive snapshots shown in B.

A large amount of experimental evidence suggests that many animal species encode a geometric layout of the surrounding space (***Cheng and Newcombe, 2005***; ***O'Keefe and Burgess, 1996***; ***Gouteux et al., 2001***; ***Krupic et al., 2015***; ***Keinath et al., 2017***; ***Bécu et al., 2019***). Computational models of spatial representation in rodents link this sensitivity to geometry with a postulated ability of the animal to estimate distances to surrounding walls (***Hartley et al., 2000***) or to observe panoramic visual snapshots of surrounding space (***Cheung et al., 2008***; ***Sheynikhovich et al., 2009***), and rely on a wide rodent visual field (320°). That the width of visual field plays a role in geometric processing in humans was demonstrated in the study by ***Sturz et al.*** (***2013***), in which limiting visual field to 50° impaired performance in a geometry-dependent navigation task, compared to a control group. We thus studied whether activities of egocentric and allocentric neurons in the model encode information about the geometry of the environment and whether snapshot accumulation over time plays a role in this process.

To do this, we run the model to uniformly explore a square environment and we stored population rate vectors of the egocentric-visual and allocentric-parietal populations at successive time points during exploration. More specifically, for the egocentric population, each population vector

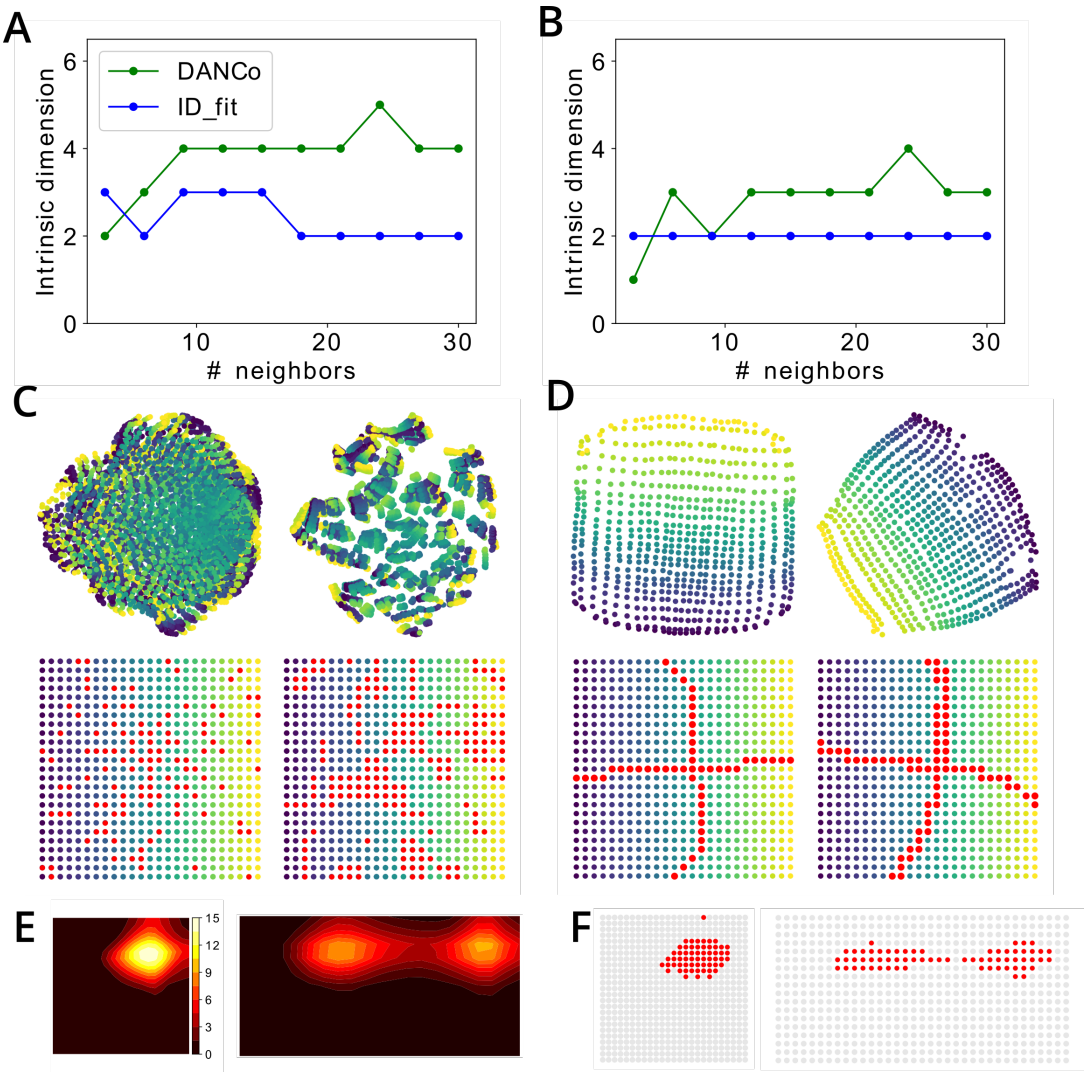

Figure 5. Representation of spatial relations by egocentric (occipito-parietal) and allocentric (parieto-retrosplenial) visual neurons. A,B. Estimation of intrinsic dimensionality of the set of population vectors in the egocentric (A) and allocentric (B) populations by two different state-of-the-art methods (DANCo and ID_fit). C,D. Top: Projection of the population vector manifolds onto a two-dimensional plane using Isomap (left) and t-SNE (right) algorithms. Color gradient from yellow to blue corresponds to the position at which the corresponding population vector was observed, as shown in the Bottom row. Red dots show population vectors that lie close to the principal axes of the 2D manifold of the principal space. C and D show population vectors of the egocentric and allocentric neuronal populations, respectively. E. An example of the receptive field of one hippocampal neuron after learning the environment before (left) and after (right) extension of the environment along it horizontal axis. F. For the same neuron as in E, red dots show locations in the environment where this neurons is winner in the WTA learning scheme.

corresponded to population activities evoked by the presentation of a single visual snapshot. In contrast, for the allocentric population, each population vector corresponded to a panoramic snapshot obtained by accumulating several successive snapshots during head rotations (see Methods). The visual information content was identical in two sets of population vectors as they were collected during the same exploration trial. Population vectors in each set can be considered as data points in a high-dimensional space of corresponding neural activities. These points are expected to belong to a two-dimensional manifold in this space, since during exploration the model animal moves in a 2D spatial plane. The analysis of the intrinsic dimensionality of both sets indeed shows that it is about 2 (***Figure 5***A,B). We then applied two different manifold visualisation techniques to see whether the

shape of manifold reflects the environment shape (see Methods). We found that when applied to population vectors of the egocentric population, the structure of the manifold did not reflect the layout of the environment (***Figure 5***C). In contrast, allocentric population activities reliably preserved geometric information in the spatial organization of the manifold (***Figure 5***D). Moreover principal axes of the manifold corresponded to the principal axes of the underlying environment only for the population vectors of the allocentric population (bottom row of ***Figure 5***C,D). The extraction of principal axes of an experimental space has been proposed to underlie spatial decision making in several experimental paradigms, including data from humans (***Gallistel, 1990***; ***Cheng and Gallistel, 2005***; ***Sturz et al., 2011***).

STDP in the connections between the parietal and hippocampal neurons ensures that allocentric spatial views are stored in memory, while lateral inhibition in the hippocampal layer implements a competition such that different hippocampal cells become selective to different localized regions of the visuospatial manifold, which, by virtue of the coherent mapping on the real space, correspond to spatial receptive fields (***Figure 5***E). When the geometry of the environment is modified, but the memorised allocentric representation remains the same, modeled hippocampal cells express corresponding modifications of their receptive fields (***Figure 5***E,F), potentially providing a purely sensory basis for the effects of geometric manipulations observed in rats (***O'Keefe and Burgess, 1996***) and humans (***Hartley et al., 2004***). These simulations show how the egocentric-allocentric conversion and short-term memory along the modeled dorsal visual pathway can be instrumental in structuring the hippocampal input according to the geometric properties of the surrounding space that were repeatedly shown to affect human navigation (***Hermer and Spelke, 1994***; ***Bécu et al., 2019***).

Visual responses of hippocampal neurons reflect learning of visual stimuli

The hippocampal memory network is thought to support a large spectrum of memory-based behaviors, and therefore its basic properties should manifest themselves in situations other than navigation. In particular, plasticity and competition, which are proposed to mediate fast hippocampal learning of visual information in our model, occur not only during navigation but also in a passive image viewing paradigm. In the next simulation inspired by the experiment of ***Jutras and Buffalo*** (***2010a***) we used the stationary model to learn a set of 100 novel images presented in a quick succession (see Methods) and recorded activities of modeled hippocampal neurons. In response to the presented stimuli, some neurons increased their firing rates as a result of STDP (winning neurons), while the rest of the neurons were inhibited (***Figure 6***A). Even though only a few neurons won the competition for each particular stimulus, some neurons were selective to a larger number of stimuli than others (***Figure 6***C,D). Therefore, stimulus-averaged firing rates of different neurons expressed either a decrease in the average firing rate (for neurons that were never winners), no change in the average rate (for neurons that were winners for a relatively small number of stimuli), or an increase in the average rate (for neurons that were winners for a relatively high number of stimuli, ***Figure 6***B). There was a larger number of neurons expressed decreased firing rates or no change, than those that increased their average rate (***Figure 6***D).

Under the assumption that a novelty-detection mechanism (assumed to reside in the hippocampus or elsewhere, but not modeled here) prevents hippocampal firing in response to a repeated stimuli, these results are in accord with the data from a number of studies showing that different subsets of recorded hippocampal neurons either decreased, showed no changes, or increased their activity in response to the presentation of a novel stimulus (***Jutras and Buffalo, 2010a***; ***Rutishauser et al., 2006***; ***Viskontas et al., 2006***). In these studies of the role of novelty in hippocampal processing, stimulus-averaged elevation of neural activity was considered as an indication of an abstract (i.e. independent of stimulus identity) novelty processing in the hippocampus (***Jutras and Buffalo, 2010a***; ***Rutishauser et al., 2006***). It is unclear how such an abstract representation of novelty can be reconciled with the role of the hippocampus in navigation. In contrast, our simulation results suggest that elevation or depression of stimulus-averaged firing rate in a neuron may be related to

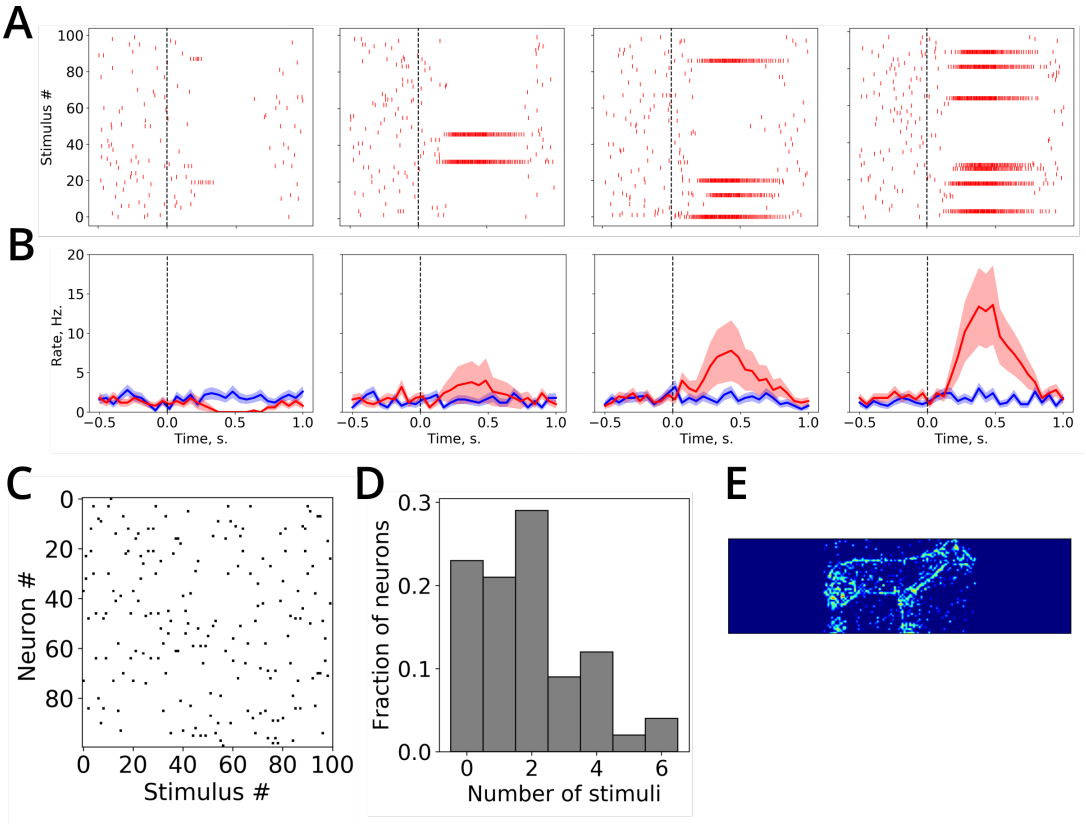

Figure 6. Visual responses of modeled hippocampal neurons. A. Spike raster plots for four example neurons in response to presented visual stimuli. B. Stimulus-averaged firing rates of neurons in A (mean ± SEM shown in red), compared to baseline firing rates (shown in blue). The dashed vertical line represents the stimulus onset. C. Black dots correspond to winner neurons among all other neurons (vertical axis) for each of the presented stimuli (horizontal axis). D. The histogram shows the distribution of neurons with respect to the number of stimuli for which they are winners. E. An example of the weight matrix of a hippocampal neuron after learning.

the number of stimuli for which this neuron is winner.

Top-down hippocampal input in spatial reorientation and memory-based search

The population of the hippocampal neurons in the model represents the neural storage of (potentially highly processed) visual information aligned with an allocentric directional frame by the coordinate transformation network. In this section we show how this neural storage can affect two types of behavior: (i) determination of position and orientation when a disoriented monkey is placed into a familiar environment (***Gouteux et al., 2001***); and (ii) memory-guided visual target search in an image viewing paradigm (***Fiehler et al., 2014***). While these two tasks may seem unrelated, we propose that the same neural process, namely a reorientation of the head-direction network based on the comparison between the newly obtained visual information and the contents of the hippocampal allocentric storage, underlies behavioral decisions in these tasks.

Spatial reorientation

In a series of reorientation experiments with monkeys, ***Gouteux et al.*** (***2001***) have shown that these animals relied on both the geometric information (given by the three-dimensional layout of the rectangular experimental space) and non-geometric cues (e.g., landmark objects placed near the walls or corners of the recording chamber). The authors paid specific attention to the influence of landmark size on reorientation behavior. When small objects were placed near one of the walls or in the corners of the room, the monkeys did not use these cues to reorient, and their search pattern was determined based only on the geometric information. Importantly, this was

not because the monkeys did not notice the landmarks, since they performed exploratory actions towards them (looked at or touched them). Once the landmark size was increased however, the monkeys successfully used them for reorientation independently of their location and number.

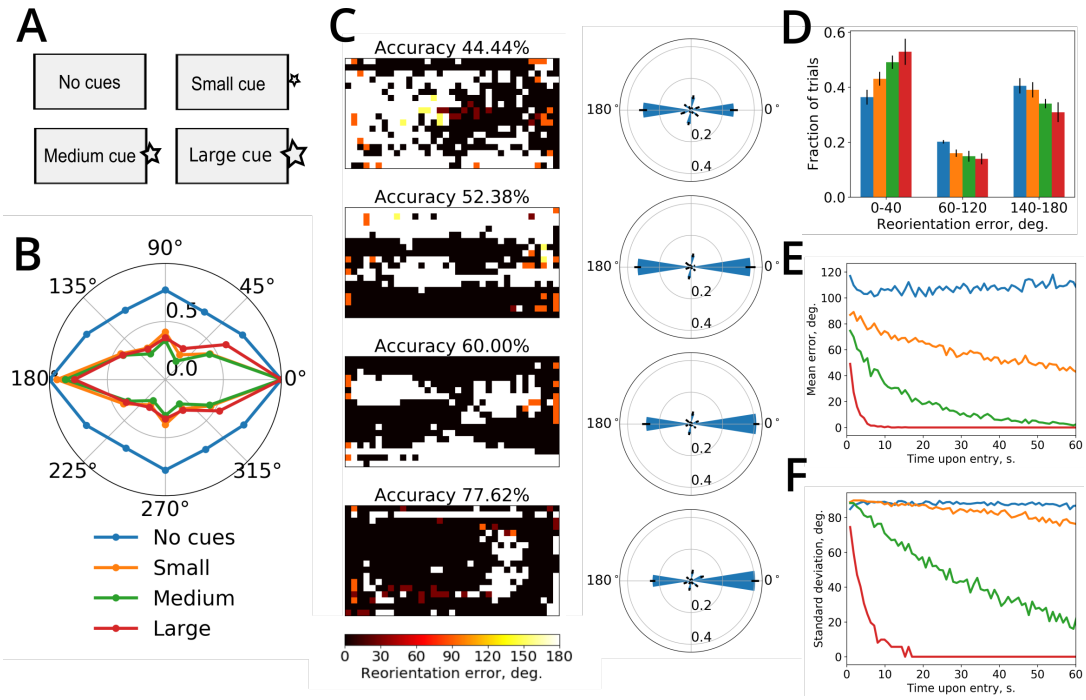

Figure 7. Simulation of the reorientation experiment. A. The experimental environment was a rectangular room (represented by the gray rectangle). The same reorientation simulation was run in four conditions: no visual cues apart from walls of the room, or 1 visual cue at three different sizes (small, medium, large). B. Polar plot of the mean activity of the reorientation network when the simulated animal was placed in various locations in the room. Dots mark the preferred locations of the reorientation N_{re} neurons Colors from blue to red represent 4 experimental conditions. C. Rows from top to bottom correspond to experimental conditions as in A. Left: Reorientation maps show, for each location in the room, the reorientation error committed by the model after seeing only the first visual snapshot from that location (at a randomly chosen head orientation). The pixel color from black to white codes for the absolute value of the reorientation error from 0 to π. Right: polar histograms of reorientation errors (±SD), averaged over 9 random orientations at each location. D. Bar plot shows the distribution of the absolute reorientation errors (±SD) among the approximately correct orientation (0-40°), rotational error (140-180°) and other directions. E,F. Reorientation error mean (E) and its standard deviation (D) when progressively more snapshots were used for reorientation. Color code for D,E,F as shown in B.

To simulate these data, we tested the model in four reorientation tasks in a virtual three-dimensional rectangular room. In these tasks, either no landmark cues were present in the room, or one visual landmark of three different sizes was placed in the middle of one of the walls (***Figure 7***A). Each task comprised an exploration phase, during which the model randomly explored the environment, and a reorientation phase. In the reorientation phase the model was initialized with a random heading direction and placed back into the environment learned during the exploration phase at a random location. The performance of the model was assessed from the accuracy of reorientation: we assume that the animal will navigate to the correct corner if it has correctly estimated its initial heading, whereas it will make a navigation error if the reorientation error is high.

Once the information from the initial view reached the hippocampus upon the reentry to the environment, the activity of the reorientation network signalled the orientation error (***Figure 7***B). This error represented the discrepancy between the initial heading direction and the heading direction most consistent with the allocentric information stored in the projections from the place

cells to the reorientation network. The asymmetric shape of the polar plot reflects the influence of the environment's geometric layout on reorientation: for the no-cue condition, the network activity peaked at the correct ($0°$) and its rotationally opposite ($180°$) orientations with an identical average amplitude. When the visual cue was present, its size determined the difference between the activity peaks. Therefore, when reorientation was performed from different locations in the environment (based only on the first view taken), the accuracy, measured as the percentage of locations with a correctly determined orientation, was about 50% in the no-cue condition and raised to about 77% in the large-cue condition (***Figure 7***C, left column). Reorientation maps (***Figure 7***C, right column) suggest that depending on the position of the orienting cue in the room, some locations in the environment provide better visual information for reorientation than others (shown by white areas in the maps). The histograms of orientation errors (***Figure 7***C, right column, and ***Figure 7***D) show that, on average, a larger visual landmark provides a much better reorienting cue than a small one, for which a similar number of correct decisions and rotational errors was observed (***Figure 7***D). This is due to the fact that orientation is determined essentially by comparing the egocentric view from the initial position with allocentric views stored in synaptic memory, without any explicit landmark identification process. Therefore, influence of small visual cues becomes negligible with respect to gross visual features of the surrounding space (corners, shapes of the walls, etc.). These results are consistent with the hypothesis that reorientation is a fast, bottom-up process based on low-level visual information (***Sheynikhovich et al., 2009***). Learning landmark identities and their spatial relation to goals can be added by subsequent learning, but may not be taken into account unless their are sufficiently salient compared to other (e.g. geometric) cues present in the environment (***Cheng, 1986***).

So far the reorientation performance was assessed based only on the first view taken. The reorientation performance is likely to increase if the animal is allowed to accumulated visual information from successive views taken in the same location at different orientations or at different locations, e.g. during initial movements through the environment. This is what happens in the model, since increasing the number of snapshots that are used for reorientation improved its accuracy (***Figure 7***E,F). In this case we placed the simulated animal at 60 successive positions, while at each position the animal rotated its head to obtain a corresponding panoramic view. The activity of the reorientation network was calculated as a sum of its activities after each successive view. When a large cue was present, the simulated animal obtained an accurate orientation estimate after visiting about 10 successive locations. In contrast, the mean error and standard deviation of reorientation were decreasing much slower for smaller sized landmarks. Thus, our model describes a neural mechanism for spatial reorientation which relies on an allocentric visual information stored in the hippocampal network. This allocentric information feeds into a head-direction-like network, assumed to reside in the retrosplenial cortex, that signals reorientation error and affect the sense of direction via its input to the head-direction system if the brain (***Taube, 2007***). In addition to providing a mechanistic basis for the reorientation process, which is a necessary part of navigational behavior and whose existence is assumed (either implicitly or explicitly) in a number of computational models of navigation, this model proposes how reorientation can be performed continuously, i.e. during ongoing spatial behavior.

Memory-based visual search

To illustrate a potential role of the stored hippocampal representation in memory-based visual tasks, we simulated the study of ***Fiehler et al.*** (***2014***). In this study, head-fixed human subjects remembered a visual scene with 6 objects on a table, presented on a computer screen (***Figure 8***A, top). This encoding phase was followed by 2-s. delay (uniform gray image), and then the subjects were presented with a modified scene in which one of the objects was missing (the target object) and either 1, 3 or 5 other objects displaced horizontally (***Figure 8***A, bottom). The subjects were required to point to the remembered location of the missing object. If the subjects had used only an egocentric information (i.e. remembered object position with respect to the head), then

their performance would have been independent from the displaced objects. The results of this experiment demonstrated in contrast that pointing performance was influenced by the non-target objects, such that shifting a higher number of them induced a larger pointing error. Even though the pointing error was always made in the direction of the object displacement in the image, the size of the error only partially accounted for the veridical displacement of the objects. These data suggest that human subjects combine allocentric (i.e. based on the information from the environment, in this case represented by the visual features associates with displaced objects) and egocentric (i.e. based on the memory of an egocentric location of the target object) information during memory-based search (***Fiehler et al., 2014***). The neural mechanism of this allocentric correction of the egocentric memory is unknown.

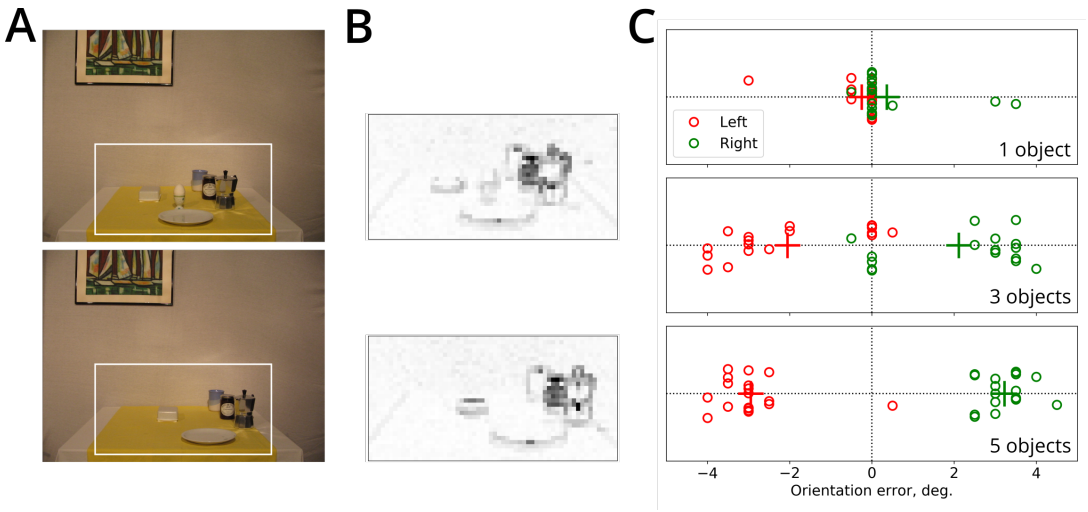

Figure 8. A. An example of the remembered (top) and test (bottom) images. In this example, the target object is the egg and 5 non-target objects were shifted to the right in the test image, compared to the encoded image. The white rectangle denotes the part of the image that was provided as input to the network. It corresponds to the part of the image most fixated by the subjects in the experiment. B. Mean firing rates of the egocentric neurons in the model for the encoded and test images shown in A. C. Orientation errors induced in the model by the presentation of the test images with 1 (top), 3 (middle) and 5 (bottom) displaced objects. Horizontal position of each dot corresponds to the maximal activity peak of the reorientation network. Different dots represent different sets of objects in the image dataset. Leftward and rightward displacements are shown in red and green, respectively. Crosses mark the mean displacement value per group. Random jitter along the vertical axis is added for clarity.

We hypothesized that the influence of allocentric image information observed in this experiment arises as a result of a slight misorientation of the head direction network due to the apparent shift of visual features caused by the object displacement in the attended area of the image. In order to demonstrate this effect, we first presented to the model an image of a control scene with all 6 objects (see ***Figure 8***A, top, for an example). We used, with permission, the same image data set that was used in the experimental study. As input to the network we only used the part of the image near the objects, because in the experiment is was fixated most of the time and because of the evidence that displacement of objects outside of this area had no influence on reaching performance (***Fiehler et al., 2014***). The network converted the visual input of the egocentric layer (***Figure 8***B) to an allocentric representation according to the actual head direction (set to 0°), which was stored in the synapses between the parieto-retrosplenial output cells and hippocampal cells as before. In this simulation we ignored competition effects, since it was not required to remember multiple images. Second, after the first scene was learned, an image of the scene with one object missing and either 1, 3 or 5 objects displaced (see ***Figure 8***B, bottom) was presented to the model. The orientation error caused by the object displacement can then be read directly from the activity of the reorientation network (***Figure 8***C). As in the experiment, the number of displaced objects affected the amount of

allocentric correction. Since in the test images the displaced objects correspond only to a subset of all visual features, the mean correction only partially account for the object displacement. Thus, as in the case of spatial reorientation, the influence of the allocentric information (in this case represented by low-level features of the presented image) is caused by the comparison between the stored allocentric and incoming allocentric views, and the resulting activity of the reorientation network that calibrates the head direction signal.

Discussion

The presented model focuses on the dorsal visual pathway for information processing, generally thought to provide contextual or "where" information to memory structures in the MTL, by contrast to the ventral pathway mediating the processing of object/item representations or "what" information (***Goodale and Milner, 1992***; ***Kravitz et al., 2011***). The two pathways converge to the hippocampus where both types of information are combined to form the episodic memories. Outputs of hippocampal processing go back to neocortical areas from which the input was originated. In both spatial (e.g. spontaneous novelty exploration) and non-spatial (recollection/familiarity) experimental paradigms the dorsal pathway has been implicated in the recollection of contextual information (e.g. the scene or location where an item was observed) and not in remembering the object identity (see ***Eichenbaum et al., 2007***, for review). These proposals go in line with general properties of neural activities along the dorsal pathway such as PHC and RCS. In particular, fMRI studies that both RSC and PHC are activated by scene processing, with a part of PHC responding equally strongly to images of spatial layouts with or without objects (***Epstein and Kanwisher, 1998***; ***Epstein, 2008***). RSC was shown to be more strongly implicated in recollection than familiarity (***Epstein, 2008***) and is proposed to play a specific role in encoding spatial and directional characteristic of landmarks and their stability *independent of their identity* (***Mitchell et al., 2018***).

In the present work, the selectivity to scenes and spatial layouts, as opposed to objects, during spatial navigation is modeled simply as sensitivity to views (i.e. the total contents of the animal's visual field at one moment in time, usually acquired across multiple fixations, potentially associated with accompanying head movements in natural conditions). Indeed, spatial layout information is often available from a low-frequency representation of a view (***Kauffmann et al., 2015***, but see ***Rajimehr et al., 2011***), whereas object representations take up a much smaller portion of a view and usually require high-spatial frequency analysis at a localized part of the image during visual fixation. In our simple model, we represented the contents of a view by a retinotopic-like grid of orientation-sensitive filter responses at just a few spatial frequencies, but a much more complex visual processing can be "inserted" between our input visual layer and the parietal transformation circuit (involving e.g., extraction of salience maps, depth processing, contour extraction, etc). The coordinate-transformation circuit and the rest of the model are agnostic about the nature of features provided to them as input, as long as these features are given in a retinotopic-like head-fixed frame and take up the whole visual field. This last requirement excludes object processing, assumed to be done in parallel in the ventral stream, since object representations are view-independent and assume translation invariance over the visual field (***Serre et al., 2005***). The relative (i.e. size dependent) sensitivity to objects in our model (see "Spatial reorientation") arises from the fact that large, distal and stable objects (or landmarks) that make up a large portion of a view are considered as part of the layout, and not as identified objects/landmarks. In contrast, relatively small objects, landmarks, or a high-frequency contents of other small localized portions of a view exert contribute only weakly to the overall visual representation. Indeed, they are often overshadowed by gross visual features present in views, such as corners, walls, and other large-scale visual structures during comparison of new and remembered view-based representations (***Bécu et al., 2019***).

Our model can thus be considered as a model of encoding of contextual information, as opposed to object-related one, and the notion of context is well defined: it is the visual information present in the set of topographically-organized features present in a set of views (that could comprise only one element) and stored in memory after the acquisition phase of a task. This notion of context can

be extended to a non-spatial setting (see "Memory-based visual search"): topographically-organized image features present in attended part of the screen and stored in memory provide contextual information with respect to any object-related information stored from the scene (such as the identities of the objects in this experiment). In the absence of reliable object-related information (such as the missing target object), contextual information can be used to drive behavior. The important piece of information that is present in topographic representation of a scene, but is absent in object-related memory, is spatial location. Indeed, one can assign position information within the topographic representation of a view (with respect to an allocentric directional frame, or with respect to the other features in the view). Therefore, (allocentric) view-based contextual representations can serve as a basis for remembering spatial and directional characteristics of objects or landmarks independent of their identity. Spatial locations in such a contextual representation can serve as "place holders" for specific object/landmark information extracted and stored in the ventral visual stream, or as "pointers" to this information. Such a notion of contextual information is well in line with proposed role of the PHC and RSC in landmark processing (***Epstein, 2008***; ***Mitchell et al., 2018***).

While the existence of view-based representations in human spatial memory is well established (***Shelton and McNamara, 1997***; ***Diwadkar and McNamara, 1997***; ***Christou and Bülthoff, 1999***; ***Garsoffky et al., 2002***; ***Burgess, 2006***), the existence of a spatiotopic representation of the surrounding visual space is more controversial. Some proposals reject the existence of such a representation (***O'Regan, 1992***), some suggest that only a limited number attended features survive beyond one fixation (***Rensink, 2000***), and some suggest that a feature-rich representation is constructed by accumulating information over time (see ***Tatler and Land, 2011***, for review). For example, some experimental evidence in favor of the latter view comes from studies showing that visual search can be directed to remembered locations in a panoramic scenes and that visual saccades can be programmed to reach previously observed targets outside of the current viewfield (***Land et al., 1999***; ***Oliva et al., 2004***). These and similar data suggest the existence of a quasi-panoramic representation of surrounding visual cues, accessible for the planning of eye movements, i.e. most likely topographic with respect to the visual space (***Golomb et al., 2011***; ***Park et al., 2007***; ***Melcher and Morrone, 2015***; ***Robertson et al., 2016***). While both egocentric and allocentric representations are stored in memory, they are converted to an egocentric frame whenever possible (***Chen et al., 2011***). By linking such a panoramic representation with its potential utility for spatial memory and the well known role of the MTL in the storage of allocentric memories, we postulated the existence of an allocentric, visually topographic representation of the surrounding space in the parieto-retrosplenial circuit.

Whereas the allocentric representation in our model is purely visual, the possibility that it could be multisensory can not be excluded (***Newell et al., 2005***). ***Loomis et al., 2013*** defined a similar representation of surrounding 3D space as a "spatial image" with the following properties: (i) it can be updated during movement with the eye closed; (ii) it exists in all directions; (iii) the information from all sensory modalities converge onto a common, "amodal", spatial image. While our model is directly consistent with the second property, the third one can be implemented by converting spatial locations of egocentric sensory signals at different modalities (e.g. haptic or auditory) into the common allocentric framework. These locations (or placeholders) can then be linked to the detailed representations of sensory experience in sensory-specific areas of the cortex, similarly to the putative links between landmark locations and their high-frequency contents discussed above. The first property can be assured by backward projections from the hippocampus to the allocentric layer (not included in the model), by a mechanism previously proposed to support spatial imagery (***Byrne et al., 2007***). One obvious candidate for the potential biological locus of the panoramic visual representation is the PPC, since spatiotopic neuronal receptive fields were observed in this area (***Snyder et al., 1998***; ***Fairhall et al., 2017***). The parahippocampal place area, a scene-selective subdivision of the PHC, while not sensitive to the images of the same scene from different viewpoints (***Epstein, 2008***), can integrate visual information across saccades to form a representation of a larger scene (***Golomb et al., 2011***). Finally, RSC and occipital place area were

recently shown to mediate the memory of panoramic visual representations (***Robertson et al., 2016***).

There are two key differences between our model and a previous influential model of spatial memory and imagery (***Becker and Burgess, 2001***; ***Byrne et al., 2007***, see also ***Bicanski and Burgess, 2016***). First, our model postulates the existence of a quasi-panoramic representation of surrounding visual space, in topographic visual coordinates, as emerging experimental evidence suggests (***Melcher and Morrone, 2015***; ***Robertson et al., 2016***). We propose that such a representation *(i)* is accumulated from successive views using short-term memory; *(ii)* can be used for planning of eye movements during natural behavior; *(iii)* serves for the storage of object/landmark position and orientation information. In our model, the reference frame for this panoramic representation is allocentric, and only a portion of it, corresponding to the current view field, is explicitly converted to an egocentric visual representation (equivalent to the "parietal window" of ***Byrne et al., 2007***). Second, our model proposes a mechanism of fast bottom-up view-based reorientation of the head direction system that was either absent (***Byrne et al., 2007***) or relied on the presence of conspicuous landmarks linked directly to head direction cells (***Bicanski and Burgess, 2016***). A number of reorientation studies mentioned earlier suggest that this neural process is independent from landmark identities and can be performed in the absence of point-like landmarks. The mechanism we use relies on weight sharing and as such is not, at its present implementation, biologically realistic. The concept of weight sharing has been critical for recent successes of brain-inspired neural networks and is widely used in models of biological networks of visual processing (e.g. ***Serre et al., 2005***; ***Masquelier and Thorpe, 2007***; ***Bartunov et al., 2018***). One possible implementation of our proposed reorientation mechanism would require mental rotation of the stored allocentric representations, while freezing the actual egocentric view in the input layer. Such an implementation would make the model significantly more complex, without changing the underlying computation.

To summarize, the model presented in this work explored the nature of visual representations in the parietal-medial temporal pathway for visuospatial processing and contributed to the open question of the link between visual and memory structures in primates. It proposes that a single, potentially multisensory, representation of surrounding environment is constructed by time-integrated sensory snapshots. This putative representation provides a 3D coordinate space within which positions of localized sensory events can be encoded and which can serve as basis for eye-movement generation in natural conditions. This model thus provides a conceptual framework for linking oculomotor behavior, visual and spatial memory.

Acknowledgments

Funding: This research was supported by ANR – Essilor SilverSight Chair ANR-14-CHIN-0001.

References

Andersen R, Snyder LHL, Li CSC, Stricanne B. Coordinate transformations in the representation of spatial information. Curr Opin Neurobiol. 1993; 3:171–176. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve{&}db=PubMed{&}dopt=Citation{&}list{_}uids=8513228{%}5Cnhttp://www.sciencedirect.com/science/article/pii/095943889390206E, doi: 8513228.

Avillac M, Denève S, Olivier E, Pouget A, Duhamel JR. Reference frames for representing visual and tactile locations in parietal cortex. Nat Neurosci. 2005 jul; 8(7):941–9. http://www.ncbi.nlm.nih.gov/pubmed/15951810, doi: 10.1038/nn1480.

Bachevalier J, Nemanic S, Alvarado MC. The influence of context on recognition memory in monkeys: Effects of hippocampal, parahippocampal and perirhinal lesions. Behav Brain Res. 2015 may; 285:89–98. https://www.sciencedirect.com/science/article/pii/S0166432814004501, doi: 10.1016/J.BBR.2014.07.010.

Bartunov S, Santoro A, Richards B, Marris L, Hinton GE, Lillicrap T. Assessing the Scalability of Biologically-Motivated Deep Learning Algorithms and Architectures. In: S B, Wallach H, Larochelle H, Grauman K, Cesa-Bianchi N, Garnett R, editors. *Adv. Neural Inf. Process. Syst. 31*; 2018. p. 9368–9378. http://papers.nips.cc/paper/8148-assessing-the-scalability-of-biologically-motivated-deep-learning-algorithms-and-architectures.

814 **Becker S**, Burgess N. Modelling spatial recall, mental imagery and neglect. Adv Neural Inf Process Syst. 2001;
815 13. http://discovery.ucl.ac.uk/4130/.

816 **Bécu M**, Sheynikhovich D, Tatur G, Agathos C, Bologna LL, Sahel JA, Arleo A. Age-related preference for geometric
817 spatial cues during real-world navigation. Nat Hum Behav. 2019; p. In press.

818 **Bicanski A**, Burgess N. Environmental Anchoring of Head Direction in a Computational Model of Retrosplenial
819 Cortex. J Neurosci. 2016; 36(46):11601–11618. doi: 10.1523/JNEUROSCI.0516-16.2016.

820 **Brandt SA**, Stark LW. Spontaneous Eye Movements During Visual Imagery Reflect the Content of the Visual
821 Scene. J Cogn Neurosci. 1997 jan; 9(1):27–38. http://www.mitpressjournals.org/doi/10.1162/jocn.1997.9.1.27,
822 doi: 10.1162/jocn.1997.9.1.27.

823 **Brotchie PR**, Andersen RA, Snyder LH, Goodman SJ. Head position signals used by parietal neurons to encode
824 locations of visual stimuli. Nature. 1995 may; 375(6528):232–235. http://www.nature.com/articles/375232a0,
825 doi: 10.1038/375232a0.

826 **Burgess N**. Spatial memory: how egocentric and allocentric combine. Trends Cogn Sci. 2006; 10(12):551–557.

827 **Burgess N**. Spatial cognition and the brain. Ann N Y Acad Sci. 2008 mar; 1124:77–97. http://www.ncbi.nlm.nih.
828 gov/pubmed/18400925, doi: 10.1196/annals.1440.002.

829 **Burgess N**. The 2014 Nobel Prize in Physiology or Medicine: A Spatial Model for Cognitive Neuroscience.
830 Neuron. 2014 dec; 84(6):1120–1125. https://www.sciencedirect.com/science/article/pii/S0896627314010903,
831 doi: 10.1016/J.NEURON.2014.12.009.

832 **Busch NA**, VanRullen R. Spontaneous EEG oscillations reveal periodic sampling of visual attention. Proc
833 Natl Acad Sci U S A. 2010 sep; 107(37):16048–53. http://www.ncbi.nlm.nih.gov/pubmed/20805482http:
834 //www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC2941320, doi: 10.1073/pnas.1004801107.

835 **Byrne P**, Becker S, Burgess N. Remembering the past and imagining the future: A neural model of spatial
836 memory and imagery. Psychol Rev. 2007; 114(2):340–375.

837 **Ceruti C**, Bassis S, Rozza A, Lombardi G, Casiraghi E, Campadelli P. DANCo: An intrinsic dimensionality estimator
838 exploiting angle and norm concentration. Pattern Recognit. 2014 aug; 47(8):2569–2581. https://www.
839 sciencedirect.com/science/article/pii/S003132031400065X, doi: 10.1016/J.PATCOG.2014.02.013.

840 **Chafee MV**, Averbeck BB, Crowe DA. Representing Spatial Relationships in Posterior Parietal Cortex: Single
841 Neurons Code Object-Referenced Position. Cereb Cortex. 2007 dec; 17(12):2914–2932. https://academic.oup.
842 com/cercor/article-lookup/doi/10.1093/cercor/bhm017, doi: 10.1093/cercor/bhm017.

843 **Chen Y**, Byrne P, Crawford JD. Time course of allocentric decay, egocentric decay, and allocentric-to-egocentric
844 conversion in memory-guided reach. Neuropsychologia. 2011 jan; 49(1):49–60. https://linkinghub.elsevier.
845 com/retrieve/pii/S0028393210004689, doi: 10.1016/j.neuropsychologia.2010.10.031.

846 **Cheng K**. A purely geometric module in the rat's spatial representation. Cognition. 1986; 23(2):149–178.

847 **Cheng K**, Gallistel CR. Shape parameters explain data from spatial transformations: comment on Pearce et
848 al. (2004) and Tommasi & Polli (2004). J Exp Psychol Anim Behav Process. 2005 apr; 31(2):254–9; discussion
849 260–1.

850 **Cheng K**, Newcombe NS. Is there a geometric module for spatial orientation? Squaring theory and evidence.
851 Psychon Bull Rev. 2005; 12(1):1–23.

852 **Cheung A**, Stürzl W, Zeil J, Cheng K. The Information Content of Panoramic Images II: The Rotational Errors
853 and the Similarity of Views in Rectangular Experimental Arenas. J Exp Psychol Anim Behav Process. 2008;
854 34(1):15–30.

855 **Christou CG**, Bülthoff HH. View dependence in scene recognition after active learning. Mem Cognit. 1999 nov;
856 27(6):996–1007. http://link.springer.com/10.3758/BF03201230, doi: 10.3758/BF03201230.

857 **Crutcher MD**, Calhoun-Haney R, Manzanares CM, Lah JJ, Levey AI, Zola SM. Eye Tracking During a Visual Paired
858 Comparison Task as a Predictor of Early Dementia. Am J Alzheimer's Dis Other Dementiasr. 2009 jun; 24(3):258–
859 266. http://journals.sagepub.com/doi/10.1177/1533317509332093, doi: 10.1177/1533317509332093.

860 **Diwadkar VA**, McNamara TP. Viewpoint Dependence in Scene Recognition. Psychol Sci. 1997 jul;
861 8(4):302–307. http://journals.sagepub.com/doi/10.1111/j.1467-9280.1997.tb00442.x, doi: 10.1111/j.1467-
862 9280.1997.tb00442.x.

Duhamel JR, Bremmer F, Ben Hamed S, Graf W. Spatial invariance of visual receptive fields in parietal cortex neurons. Nature. 1997; 389(6653):845–8. doi: 10.1038/39865.

Durstewitz D, Seamans JK, Sejnowski TJ. Dopamine-Mediated Stabilization of Delay-Period Activity in a Network Model of Prefrontal Cortex. J Neurophysiol. 2000; 83(3):1733–1750.

Eichenbaum H, Yonelinas AP, Ranganath C. The Medial Temporal Lobe and Recognition Memory. Annu Rev Neurosci. 2007 jul; 30(1):123–152. http://www.annualreviews.org/doi/10.1146/annurev.neuro.30.051606.094328, doi: 10.1146/annurev.neuro.30.051606.094328.

Ekstrom AD, Kahana MJ, Caplan JB, Fields TA, Isham EA, Newman EL, Fried I. Cellular networks underlying human spatial navigation. Nature. 2003; 425(6954):184–188. doi: 10.1038/nature01964.

Epstein R, Kanwisher N. A cortical representation of the local visual environment. Nature. 1998 apr; 392(6676):598–601. http://www.ncbi.nlm.nih.gov/pubmed/9560155, doi: 10.1038/33402.

Epstein Ra. Parahippocampal and retrosplenial contributions to human spatial navigation. Trends Cogn Sci. 2008; 12(10):388–396. doi: 10.1016/j.tics.2008.07.004.

Epstein RA, Vass LK. Neural systems for landmark-based wayfinding in humans. Philos Trans R Soc Lond B Biol Sci. 2014 feb; 369(1635):20120533. http://rstb.royalsocietypublishing.org/content/369/1635/20120533, doi: 10.1098/rstb.2012.0533.

Fairhall SL, Schwarzbach J, Lingnau A, Van Koningsbruggen MG, Melcher D. Spatiotopic updating across saccades revealed by spatially-specific fMRI adaptation. Neuroimage. 2017 feb; 147:339–345. https://www.sciencedirect.com/science/article/pii/S1053811916306899, doi: 10.1016/J.NEUROIMAGE.2016.11.071.

Fiehler K, Wolf C, Klinghammer M, Blohm G. Integration of egocentric and allocentric information during memory-guided reaching to images of a natural environment. Front Hum Neurosci. 2014 aug; 8(August):1–12. http://www.frontiersin.org/Human{_}Neuroscience/10.3389/fnhum.2014.00636/abstract, doi: 10.3389/fnhum.2014.00636.

Gallistel CR. The organization of learning. Cambridge, MA: MIT Press; 1990.

Garsoffky B, Schwan S, Hesse FW. Viewpoint dependency in the recognition of dynamic scenes. J Exp Psychol Learn Mem Cogn. 2002 nov; 28(6):1035–50. http://www.ncbi.nlm.nih.gov/pubmed/12450330.

Gaunet F, Vidal M, Kemeny A, Berthoz A. Active, passive and snapshot exploration in a virtual environment: Influence on scene memory, reorientation and path memory. Cogn brain Res. 2001; 11(3):409–420.

Gerstner W, Kistler WM, Naud R, Paninski L. Neuronal Dynamics. Cambridge: Cambridge University Press; 2014. http://ebooks.cambridge.org/ref/id/CBO9781107447615, doi: 10.1017/CBO9781107447615.

Goldman-Rakic PS. Cellular basis of working memory. Neuron. 1995; 14(3):477–485.

Golomb JD, Albrecht AR, Park S, Chun MM. Eye movements help link different views in scene-selective cortex. Cereb Cortex. 2011 sep; 21(9):2094–102. http://cercor.oxfordjournals.org/content/21/9/2094.abstracthttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3155605{&}tool=pmcentrez{&}rendertype=abstracthttps://academic.oup.com/cercor/article-lookup/doi/10.1093/cercor/bhq292, doi: 10.1093/cercor/bhq292.

Goodale MA, Milner AD. Separate visual pathways for perception and action. Trends Neurosci. 1992; 15(1):20–25. http://www.cnbc.cmu.edu/braingroup/papers/goodale{_}milner{_}1992.pdfhttp://books.google.com/books?hl=en{&}lr={&}id=j5q0VvOGExYC{&}oi=fnd{&}pg=PA175{&}dq={%}2522the+identification+of+that+stimulus.+He{%}2522+{%}2522spatial+location+of+visual+stimuli,+only+some{%}2522+{%}2522198, doi: 10.1016/0166-2236(92)90344-8.

Gouteux S, Thinus-Blanc C, Vauclair J. Rhesus monkeys use geometric and nongeometric information during a reorientation task. J Exp Psychol Gen. 2001; 130(3):505–19.

Granata D, Carnevale V. Accurate Estimation of the Intrinsic Dimension Using Graph Distances: Unraveling the Geometric Complexity of Datasets. Sci Rep. 2016 nov; 6(1):31377. http://www.nature.com/articles/srep31377, doi: 10.1038/srep31377.

Hartley T, Burgess N, Lever C, Cacucci F, Keefe JO. Modeling place fields in terms of the cortical inputs to the hippocampus. Hippocampus. 2000; 10:369–379.

Hartley T, Trinkler I, Burgess N. Geometric determinants of human spatial memory. Cognition. 2004; 94(1):39–75. doi: 10.1016/j.cognition.2003.12.001.

Hasselmo ME, Bodelon C, Wyble BP. A proposed function for hippocampal theta rhythm: separate phases of encoding and retrieval enhance reversal of prior learning. Neural Comput. 2002 apr; 14(4):793–817.

Hayhoe MM, Shrivastava A, Mruczek R, Pelz JB. Visual memory and motor planning in a natural task. J Vis. 2003 feb; 3(1):6. http://jov.arvojournals.org/article.aspx?doi=10.1167/3.1.6, doi: 10.1167/3.1.6.

Heeger DJ. Normalisation of cell responses in cat striate cortex. Vis Neurosci. 1992; 9:181–197. doi: 10.1017/S0952523800009640.

Hermer L, Spelke ES. A geometric process for spatial reorientation in young children. Nature. 1994; 370:57–59. doi: 10.1038/370057a0.

Hoffman KL, Dragan MC, Leonard TK, Micheli C, Montefusco-Siegmund R, Valiante TA. Saccades during visual exploration align hippocampal 3–8 Hz rhythms in human and non-human primates. Front Syst Neurosci. 2013 aug; 7:43. http://journal.frontiersin.org/article/10.3389/fnsys.2013.00043/abstract, doi: 10.3389/fnsys.2013.00043.

Hori E, Tabuchi E, Matsumura N, Tamura R, Eifuku S, Endo S, Nishijo H, Ono T. Representation of place by monkey hippocampal neurons in real and virtual translocation. Hippocampus. 2003; 13(2):190–6. doi: 10.1002/hipo.10062.

Johansson R, Johansson M. Look Here, Eye Movements Play a Functional Role in Memory Retrieval. Psychol Sci. 2014 jan; 25(1):236–242. http://journals.sagepub.com/doi/10.1177/0956797613498260, doi: 10.1177/0956797613498260.

Jutras MJ, Buffalo EA. Recognition memory signals in the macaque hippocampus. Proc Natl Acad Sci. 2010; 107(1):401–406. http://www.pnas.org/cgi/doi/10.1073/pnas.0908378107, doi: 10.1073/pnas.0908378107.

Jutras MJ, Buffalo EA. Synchronous neural activity and memory formation. Curr Opin Neurobiol. 2010; 20(2):150–155. http://dx.doi.org/10.1016/j.conb.2010.02.006, doi: 10.1016/j.conb.2010.02.006.

Kauffmann L, Ramanoël S, Guyader N, Chauvin A, Peyrin C. Spatial frequency processing in scene-selective cortical regions. Neuroimage. 2015 may; 112:86–95. https://www.sciencedirect.com/science/article/pii/S1053811915001688, doi: 10.1016/J.NEUROIMAGE.2015.02.058.

Keinath AT, Julian JB, Epstein RA, Muzzio IA. Environmental Geometry Aligns the Hippocampal Map during Spatial Reorientation. Curr Biol. 2017 feb; 27(3):309–317. http://www.sciencedirect.com/science/article/pii/S0960982216314002http://linkinghub.elsevier.com/retrieve/pii/S0960982216314002, doi: 10.1016/j.cub.2016.11.046.

Kravitz DJ, Saleem KS, Baker CI, Mishkin M. A new neural framework for visuospatial processing. Nat Rev Neurosci. 2011; 12(4):217–230. http://dx.doi.org/10.1038/nrn3008, doi: 10.1167/11.11.923.

Krupic J, Bauza M, Burton S, Barry C, O'Keefe J. Grid cell symmetry is shaped by environmental geometry. Nature. 2015; 518(7538):232–235. doi: 10.1038/nature14153.

Laeng B, Bloem IM, D'Ascenzo S, Tommasi L. Scrutinizing visual images: The role of gaze in mental imagery and memory. Cognition. 2014 may; 131(2):263–283. https://www.sciencedirect.com/science/article/pii/S0010027714000043, doi: 10.1016/J.COGNITION.2014.01.003.

Land M, Mennie N, Rusted J. The Roles of Vision and Eye Movements in the Control of Activities of Daily Living. Perception. 1999 nov; 28(11):1311–1328. http://journals.sagepub.com/doi/10.1068/p2935, doi: 10.1068/p2935.

Land MF. Do we have an internal model of the outside world? Philos Trans R Soc Lond B Biol Sci. 2014; 369:20130045. http://www.ncbi.nlm.nih.gov/pubmed/24395972, doi: 10.1098/rstb.2013.0045.

Loomis JM, Klatzky RL, Giudice NA. Representing 3D Space in Working Memory: Spatial Images from Vision, Hearing, Touch, and Language. In: Lacey S, Lawson R, editors. *Multisensory Imag.* New York, NY: Springer New York; 2013.p. 131–155. https://doi.org/10.1007/978-1-4614-5879-1{_}8http://link.springer.com/10.1007/978-1-4614-5879-1{_}8, doi: 10.1007/978-1-4614-5879-1_8.

van der Maaten L, Hinton G. Visualizing Data using t-SNE. J Mach Learn Res. 2008; 9(Nov):2579–2605. http://www.jmlr.org/papers/v9/vandermaaten08a.html.

960 **Manns JR**, Stark CE, Squire LR. The visual paired-comparison task as a measure of declarative memory.
961 Proc Natl Acad Sci U S A. 2000 oct; 97(22):12375–9. http://www.ncbi.nlm.nih.gov/pubmed/11027310http:
962 //www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC17349, doi: 10.1073/pnas.220398097.

963 **Masquelier T**, Thorpe SJ. Unsupervised learning of visual features through spike timing dependent plasticity.
964 PLoS Comput Biol. 2007 feb; 3(2):e31. http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.
965 0030031, doi: 10.1371/journal.pcbi.0030031.

966 **McGregor A**, Hayward AJ, Pearce JM, Good MA. Hippocampal Lesions Disrupt Navigation Based on the Shape
967 of the Environment. Behav Neurosci. 2004 oct; 118(5):1011–1021. http://doi.apa.org/getdoi.cfm?doi=10.1037/
968 0735-7044.118.5.1011, doi: 10.1037/0735-7044.118.5.1011.

969 **McKee RD**, Squire LR. On the development of declarative memory. J Exp Psychol Learn Mem Cogn.
970 1993; 19(2):397–404. http://doi.apa.org/getdoi.cfm?doi=10.1037/0278-7393.19.2.397, doi: 10.1037/0278-
971 7393.19.2.397.

972 **Meister MLR**, Buffalo EA. Getting directions from the hippocampus: The neural connection between looking
973 and memory. Neurobiol Learn Mem. 2016; 134:135–144. doi: 10.1016/j.nlm.2015.12.004.

974 **Melcher D**, Morrone MC. Nonretinotopic visual processing in the brain. Vis Neurosci. 2015 sep;
975 32:E017. https://www.cambridge.org/core/product/identifier/S095252381500019X/type/journal{_}article,
976 doi: 10.1017/S095252381500019X.

977 **Mitchell AS**, Czajkowski R, Zhang N, Jeffery K, Nelson AJD. Retrosplenial cortex and its role in spatial cogni-
978 tion. Brain Neurosci Adv. 2018 jan; 2:239821281875709. https://www.biorxiv.org/content/early/2017/09/19/
979 190801http://journals.sagepub.com/doi/10.1177/2398212818757098, doi: 10.1177/2398212818757098.

980 **Moser EI**, Moser MB, McNaughton BL. Spatial representation in the hippocampal formation: a history. Nat
981 Neurosci. 2017 nov; 20(11):1448–1464. http://www.ncbi.nlm.nih.gov/pubmed/29073644http://www.nature.
982 com/articles/nn.4653, doi: 10.1038/nn.4653.

983 **Newell FN**, Woods AT, Mernagh M, B\"ulthoff HH. Visual, haptic and crossmodal recognition of scenes. Exp Brain
984 Res. 2005 feb; 161(2):233–242. http://link.springer.com/10.1007/s00221-004-2067-y, doi: 10.1007/s00221-
985 004-2067-y.

986 **O'Keefe J**, Burgess N. Geometric determinants of the place fields of hippocam-
987 pal neurons. Nature. 1996; 381:425–428. http://us.mc301.mail.yahoo.com/
988 mc/welcome?.gx=1{&}.tm=1278074434{&}.rand=0dilrrki4et57{#}{_}pg=showFolder;{_}ylc=
989 X3oDMTBucnBqYmU5BF9TAzM5ODMwMTAxMgRhYwNkZWxNc2dz{&}{&}filterBy={&}fid=
990 {%}2540B{%}2540Bulk{&}.rand=35722358{&}nsc{&}hash=e7f17757f01dcf1077eef51085f2bdd7{&}.jsrand=
991 8241224.

992 **Oliva A**, Wolfe JM, Arsenio HC. Panoramic Search: The Interaction of Memory and Vision in Search Through a
993 Familiar Scene. J Exp Psychol Hum Percept Perform. 2004; 30(6):1132–1146. http://doi.apa.org/getdoi.cfm?
994 doi=10.1037/0096-1523.30.6.1132, doi: 10.1037/0096-1523.30.6.1132.

995 **O'Regan JK**. Solving the "real" mysteries of visual perception: the world as an outside memory. Can J Psychol.
996 1992 sep; 46(3):461–88. http://www.ncbi.nlm.nih.gov/pubmed/1486554, doi: 10.1037/h0084327.

997 **Park S**, Intraub H, Yi DJ, Widders D, Chun MM. Beyond the Edges of a View: Boundary Extension in Human
998 Scene-Selective Visual Cortex. Neuron. 2007 apr; 54(2):335–342. https://www.sciencedirect.com/science/
999 article/pii/S0896627307002565, doi: 10.1016/J.NEURON.2007.04.006.

1000 **Pascalis O**, Hunkin NM, Bachevalier J, Mayes AR. Change in background context disrupts per-
1001 formance on visual paired comparison following hippocampal damage. Neuropsychologia. 2009
1002 aug; 47(10):2107–2113. https://www.sciencedirect.com/science/article/pii/S0028393209001626, doi:
1003 10.1016/J.NEUROPSYCHOLOGIA.2009.04.001.

1004 **Pouget A**, Deneve S, Duhamel Jr. A computational perspective on the neural basis of multisensory spatial
1005 representations. Nat Neurosci. 2002; 3(September):1–7.

1006 **Pouget A**, Sejnowski TJ. Spatial Transformations in the Parietal Cortex Using Basis Functions. J Cogn
1007 Neurosci. 1997; 9(2):222–237. http://www.mitpressjournals.org/doi/abs/10.1162/jocn.1997.9.2.222http:
1008 //www.mitpressjournals.org/doi/10.1162/jocn.1997.9.2.222, doi: 10.1162/jocn.1997.9.2.222.

Rajimehr R, Devaney KJ, Bilenko NY, Young JC, Tootell RBH. The "Parahippocampal Place Area" Responds Preferentially to High Spatial Frequencies in Humans and Monkeys. PLoS Biol. 2011 apr; 9(4):e1000608. https://dx.plos.org/10.1371/journal.pbio.1000608, doi: 10.1371/journal.pbio.1000608.

Rensink RA. The Dynamic Representation of Scenes. Vis cogn. 2000 jan; 7(1-3):17–42. http://www.tandfonline.com/doi/abs/10.1080/135062800394667, doi: 10.1080/135062800394667.

Robertson C, Hermann K, Mynick A, Kravitz D, Kanwisher N. Neural Representations Integrate the Current Field of View with the Remembered 360° Panorama in Scene-Selective Cortex. Curr Biol. 2016 sep; 26(18):2463–2468. https://www.sciencedirect.com/science/article/pii/S0960982216307539, doi: 10.1016/J.CUB.2016.07.002.

Rutishauser U, Mamelak AN, Schuman EM. Single-Trial Learning of Novel Stimuli by Individual Neurons of the Human Hippocampus-Amygdala Complex. Neuron. 2006 mar; 49(6):805–813. https://www.sciencedirect.com/science/article/pii/S0896627306001322, doi: 10.1016/J.NEURON.2006.02.015.

Salinas E, Abbott LF. Coordinate transformations in the visual system: how to generate gain fields and what to compute with them. Prog Brain Res. 2001; 130:175–90.

Salinas E, Abbott L. Transfer of coded information from sensory to motor networks. J Neurosci. 1995 oct; 15(10):6461–6474. http://www.jneurosci.org/cgi/content/abstract/15/10/6461.

Serre T, Wolf L, Poggio T. Object Recognition with Features Inspired by Visual Cortex. In: *2005 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2 IEEE; 2005. p. 994–1000. http://ieeexplore.ieee.org/document/1467551/, doi: 10.1109/CVPR.2005.254.

Shelton AL, McNamara TP. Multiple views of spatial memory. Psychon Bull Rev. 1997 mar; 4(1):102–106. http://www.springerlink.com/index/10.3758/BF03210780, doi: 10.3758/BF03210780.

Sheynikhovich D, Chavarriaga R, Strösslin T, Arleo A, Gerstner W. Is there a geometric module for spatial orientation? Insights from a rodent navigation model. Psychol Rev. 2009; 116(3):540–566. doi: 10.1037/a0016170.

Snyder LH, Grieve KL, Brotchie P, Andersen Ra. Separate body- and world-referenced representations of visual space in parietal cortex. Nature. 1998; 394(6696):887–91. doi: 10.1038/29777.

Sobotka S, Nowicka A, Ringo JL. Activity Linked to Externally Cued Saccades in Single Units Recorded From Hippocampal, Parahippocampal, and Inferotemporal Areas of Macaques. J Neurophysiol. 1997 oct; 78(4):2156–2163. https://www.physiology.org/doi/10.1152/jn.1997.78.4.2156, doi: 10.1152/jn.1997.78.4.2156.

Stimberg M, Brette R, Goodman DF. Brian 2, an intuitive and efficient neural simulator. Elife. 2019 aug; 8. https://elifesciences.org/articles/47314, doi: 10.7554/eLife.47314.

Sturz BR, Gurley T, Bodily KD. Orientation in trapezoid-shaped enclosures: implications for theoretical accounts of geometry learning. J Exp Psychol Anim Behav Process. 2011 apr; 37(2):246–53. http://www.ncbi.nlm.nih.gov/pubmed/21319918, doi: 10.1037/a0021215.

Sturz BR, Kilday ZA, Bodily KD. Does constraining field of view prevent extraction of geometric cues for humans during virtual-environment reorientation? J Exp Psychol Anim Behav Process. 2013; 39(4):390–396. http://doi.apa.org/getdoi.cfm?doi=10.1037/a0032543, doi: 10.1037/a0032543.

Tatler BW, Gilchrist ID, Rusted J. The Time Course of Abstract Visual Representation. Perception. 2003 may; 32(5):579–592. http://journals.sagepub.com/doi/10.1068/p3396, doi: 10.1068/p3396.

Tatler BW, Land MF. Vision and the representation of the surroundings in spatial memory. Philos Trans R Soc Lond B Biol Sci. 2011; 366(1564):596–610. doi: 10.1098/rstb.2010.0188.

Taube JS. The Head Direction Signal: Origins and Sensory-Motor Integration. Annu Rev Neurosci. 2007 jul; 30(1):181–207. http://www.annualreviews.org/doi/10.1146/annurev.neuro.29.051605.112854, doi: 10.1146/annurev.neuro.29.051605.112854.

Tenenbaum JB, de Silva V, Langford JC. A global geometric framework for nonlinear dimensionality reduction. Science (80-). 2000 dec; 290(5500):2319–23.

Viskontas IV, Knowlton BJ, Steinmetz PN, Fried I. Differences in Mnemonic Processing by Neurons in the Human Hippocampus and Parahippocampal Regions. J Cogn Neurosci. 2006 oct; 18(10):1654–1662. http://www.mitpressjournals.org/doi/10.1162/jocn.2006.18.10.1654, doi: 10.1162/jocn.2006.18.10.1654.

1057 **Vogt BA**, Finch DM, Olson CR. Functional Heterogeneity in Cingulate Cortex: The Anterior Executive and Posterior
1058 Evaluative Regions. Cereb Cortex. 1992 nov; 2(6):435–443. https://academic.oup.com/cercor/article-lookup/
1059 doi/10.1093/cercor/2.6.435-a, doi: 10.1093/cercor/2.6.435-a.

1060 **Wilson FAW**, Goldman-Rakic PS. Viewing preferences of rhesus monkeys related to memory for complex
1061 pictures, colours and faces. Behav Brain Res. 1994 jan; 60(1):79–89. https://www.sciencedirect.com/science/
1062 article/pii/0166432894900663, doi: 10.1016/0166-4328(94)90066-3.

1063 **Zhang K**. Representation of spatial orientation by the intrinsic dynamics of the head-direction cell ensemble: a
1064 theory. J Neurosci. 1996 mar; 16(6):2112–2126. http://www.jneurosci.org/content/16/6/2112.short.

1065 **Zipser D**, Andersen RA. A back-propagation programmed network that simulates response properties of a
1066 subset of posterior parietal neurons. Nature. 1988; 331:679–684.

1067 **Zola SM**, Squire LR, Teng E, Stefanacci L, Buffalo EA, Clark RE. Impaired recognition memory in monkeys after
1068 damage limited to the hippocampal region. J Neurosci. 2000 jan; 20(1):451–63. http://www.ncbi.nlm.nih.gov/
1069 pubmed/10627621, doi: 10.1523/JNEUROSCI.20-01-00451.2000.

1070 **Zola SM**, Manzanares CM, Clopton P, Lah JJ, Levey AI. A Behavioral Task Predicts Conversion to Mild Cognitive
1071 Impairment and Alzheimer's Disease. Am J Alzheimer's Dis Other Dementiasr. 2013 mar; 28(2):179–184.
1072 http://journals.sagepub.com/doi/10.1177/1533317512470484, doi: 10.1177/1533317512470484.