Spatial coordinate transforms linking the allocentric hippocampal and egocentric parietal primate brain systems for memory, action in space, and navigation

Edmund T Rolls

Oxford Centre for Computational Neuroscience, Oxford, UK and
University of Warwick, Department of Computer Science, Coventry, UK

Email: Edmund.Rolls@oxcns.org
Url: www.oxcns.org

Running title: Parietal cortex and hippocampus

Abstract

A theory and model of spatial coordinate transforms in the dorsal visual system through the parietal cortex that enable an interface via posterior cingulate and related retrosplenial cortex to allocentric spatial representations in the primate hippocampus is described. First, a new approach to coordinate transform learning in the brain is proposed, in which the traditional gain modulation is complemented by temporal trace rule competitive network learning. It is shown in a computational model that the new approach works much more precisely than gain modulation alone, by enabling neurons to represent the different combinations of signal and gain modulator more accurately. This understanding may have application to many brain areas where coordinate transforms are learned. Second, a set of coordinate transforms is proposed for the dorsal visual system / parietal areas that enables a representation to be formed in allocentric spatial view coordinates. The input stimulus is merely a stimulus at a given position in retinal space, and the gain modulation signals needed are eye position, head direction, and place, all of which are present in the primate brain. Neurons that encode the bearing to a landmark are involved in the coordinate transforms. Part of the importance here is that the coordinates of the allocentric view produced in this model are the same as those of spatial view cells that respond to allocentric view recorded in the primate hippocampus and parahippocampal cortex. The result is that information from the dorsal visual system can be used to update the spatial input to the hippocampus in the appropriate allocentric coordinate frame, including providing for idiothetic update to allow for self-motion. It is further shown how hippocampal spatial view cells could be useful for the transform from hippocampal allocentric coordinates to egocentric coordinates useful for actions in space and for navigation.

1. Introduction

The primate, including human, hippocampus receives inputs from the ventral visual stream about objects (the 'what') pathway (Ungerleider and Haxby, 1994)), and from the dorsal visual stream spatial representations (the 'where') pathway, as shown in Fig. 1. There is evidence that the hippocampus then can associate what and where representations in its CA3 network to form episodic memories (Kesner and Rolls, 2015). (The hypothesis was developed in a number of studies (Rolls, 1987; Rolls, 1989a; Rolls, 1989b; Treves and Rolls, 1994), and is supported by findings that hippocampal CA3 neurons in primates respond to combinations of what (object) and where (location) stimuli that need to be associated (Rolls et al., 1989; Rolls et al., 2005), and that CA3 disruption interferes with these associations in rodents (Gold and Kesner, 2005; Kesner et al., 2008; Moser and Moser, 2003; Nakazawa et al., 2002), as described in detail by Kesner and Rolls (2015).) The primate hippocampal representations are in allocentric, world-based, coordinates as shown by the responses of primate hippocampal spatial view cells, place cells, and place-dependent landmark cells (Georges-François et al., 1999; Hazama and Tamura, 2019; Kesner and Rolls, 2015; Rolls, 2018; Rolls and O'Mara, 1995; Rolls and Wirth, 2018; Wirth et al., 2017). Allocentric encoding of place was discovered in the rat hippocampus by O'Keefe and colleagues (McNaughton et al., 1983; O'Keefe, 1979; O'Keefe and Dostrovsky, 1971), and allocentric representations of the location 'out there' encoded by spatial view cells were discovered by Rolls and colleagues (Feigenbaum and Rolls, 1991; Georges-François et al., 1999; Robertson et al., 1998; Rolls et al., 1989; Rolls and O'Mara, 1995; Rolls et al., 1997a; Rolls et al., 1998; Rolls and Xiang, 2005; Rolls et al., 2005). These spatial view cells are especially suitable for primate including human visual episodic memory, of, for example, where objects are located in viewed space. Some of the information for building spatial scene representations is likely to come from the ventral visual system. In particular, inputs to the hippocampal spatial view cells may come at least in part from temporal lobe and related cortical areas that respond to scenes or parts of scenes (Kornblith et al., 2013; Nasr et al., 2011). On the other hand, the idiothetic update of hippocampal spatial view cells may involve inputs from the spatial view grid cells in the macaque medial entorhinal cortex (Killian et al., 2012; Meister and Buffalo, 2018), which in turn may be influenced by inputs from the dorsal visual system and parietal cortex as described in this paper. The mechanism for this transformation has been proposed to be competitive learning in the hippocampus (Rolls and Wirth, 2018), in a way that is analogous to the competitive learning model for the transform from rodent entorhinal place grid cells to hippocampal place cells. The primate spatial view grid cells may be related to primate hippocampal spatial view cells in that both respond when a macaque looks at different places on a screen (Feigenbaum and Rolls, 1991; Killian et al., 2012; Meister and Buffalo, 2018), and it will be interesting in future research on primate entorhinal cortex spatial view grid cells to know whether they respond to the same locations on the screen (or in space) when the macaque is translated to a different place, which is a property of primate hippocampal spatial view cells (Feigenbaum and Rolls, 1991; Georges-François et al., 1999; Robertson et al., 1998; Rolls and O'Mara, 1995; Rolls et al., 1997a; Rolls et al., 1998). In turn, the primate medial entorhinal cortex receives its inputs via the parahippocampal gyrus (areas TH and TF, Fig. 1), which in turn receives its inputs from posterior cingulate cortex, retrosplenial cortex, and related dorsal stream visual areas such as parietal area 7a (Kobayashi and Amaral, 2003) and including VIP, in which neurons that respond to signals related to self-motion are found (Bremmer et al., 2002; Chen et al., 2018; Duhamel et al., 1997; Galletti and Fattori, 2018), and LIP in which neurons with saccade-related activity are found (Graf and Andersen, 2014). The question then arises of how coordinate transforms occur in the dorsal visual system, which start in retinal coordinates in V1 which are egocentric, so that they can be useful in the hippocampal allocentric processes, including associating what with where. Signals useful for this include head direction, which is known to be represented in the primate presubiculum (Robertson et al., 1999), as well as in rodents (Cullen and Taube, 2017).

The aim of this paper is to address the issue of coordinate transforms in the dorsal visual system, in the context of how dorsal stream visual signals may interface to the hippocampal system in areas such as the retrosplenial and posterior cingulate cortex shown in Fig. 1. A new theory of the set of coordinate transforms that are performed in the primate dorsal visual system, and of how they are performed, is described, and is then tested, and the ideas further developed in a computational model.

In previous work, the importance of coordinate transforms utilizing allocentric and egocentric representations for spatial navigation in primates including humans has been described

though without a formal neuronal network theory and model (Ekstrom et al., 2017). A previous model for coordinate transforms between egocentric and allocentric coordinates holds that this is performed in the retrosplenial cortex, is described in the context of a model of spatial memory, imagery, etc., and depends on neurons such as allocentric boundary-vector cells found in rodents with consistent human fMRI evidence (Shine et al., 2019) but not shown at the neuronal level in primates, and object-vector cells, and does not model the series of stages of the highly developed primate dorsal visual system leading to the parietal cortex with coordinate transforms starting with retinal coordinates (Bicanski and Burgess, 2018; Burgess and Hartley, 2001; Byrne et al., 2007) (see Discussion). Further, the homology between the well-defined primate retrosplenial cortex of primates (Kobayashi and Amaral, 2003) and what is described as retrosplenial cortex in rodents is not clear, and there may be no posterior cingulate cortex in rodents (Vogt, 2009). Previous work has focused on the gain modulation by eye position to transform from retinal to head-centred coordinates (Pouget and Sejnowski, 1997; Salinas and Abbott, 1995; Salinas and Abbott, 1996; Salinas and Abbott, 2001; Salinas and Sejnowski, 2001). In addition. a backpropagation of error network has been trained to produce some of the types of neuron found in the parietal cortex (Zipser and Andersen, 1988), but that approach does of course not provide a theory and model of how the coordinate transforms are performed in the parietal cortex. However, this is addressed in the present research, starting with stimuli in retinal coordinates, and progressing through three stages of coordinate transform in the primate dorsal visual system to reach a representation that is in allocentric spatial view coordinates, coding for a location on space "out there", a landmark, independently of the place where the viewer is located, and of the egocentric direction or allocentric bearing of the location, e.g. a landmark, in space.

It is noted that core navigation processes and actions in space may include transformations from allocentric representations to egocentric motor commands, used for example to reach out to a remembered location of an object or to navigate through space. Indeed, the retrosplenial cortex (Kobayashi and Amaral, 2003) is implicated in navigation, spatial memory, imagery, etc., (Alexander and Nitz, 2015; Byrne et al., 2007; Epstein, 2008; Vann et al., 2009; Vedder et al., 2017), and lesions to the neocortex can produce topographical agnosia and inability to navigate (Barton, 2011; Kolb and Whishaw, 2015). In more detail, lesions restricted to the hippocampus in humans result only in slight navigation impairments in familiar environments, but rather strongly impair learning or imagining new trajectories (Bohbot and Corkin, 2007; Clark and Maguire, 2016; Maguire et al., 2016; Spiers and Maguire, 2006; Teng and Squire, 1999). In contrast, lesions in regions such as the parietal cortex or the retrosplenial cortex produce strong topographical disorientation in both familiar and new environments (Aguirre and D'Esposito, 1999; Habib and Sirigu, 1987; Kim et al., 2015; Maguire, 2001; Takahashi et al., 1997). This suggests that the core navigation processes (which may include transformations from allocentric representations to egocentric motor commands) is performed independently by neocortical (including parietal cortex) areas outside the hippocampus, which may utilize hippocampal information related to recent memories (Ekstrom et al., 2014; Miller et al., 2013; Rolls and Wirth, 2018). The ways in which useful representations could be produced for the hippocampus as outputs of the dorsal visual system via the parietal cortex in areas such as the retrosplenial cortex are part of what the theory and model introduced here address.

For the current purposes, allocentric representations are where the reference frame is the world, for example a particular location in the world, or a bearing direction. A bearing direction is a direction to a stimulus or landmark from the place where one is located. The bearing is with reference to the world, that is, is in allocentric coordinates, and is usually provided as the angle relative to North, which provides an allocentric reference frame. (Bearing direction is well known to navigators, who use the bearings of several landmarks to identify the place in the world where they are located.) The bearing direction of a landmark is different from head direction; and from the direction of motion or course travelled of the individual, vessel, etc.). Egocentric representations are where the reference frame is with respect to the head, body, etc, independently of where the organism is, or objects are, in allocentric space.

2. The theory of coordinate transforms in the dorsal visual system for use in hippocampus-related functions.

The theory for a single coordinate transform performed by a single stage of cortical processing

The theory is that each stage of coordinate transform involves two processes, a gain modulation, and then slow learning involving a short-term memory synaptic trace learning rule. The very novel part of the proposal is the use of the trace learning rule, which has been shown to be very

useful in helping to build object representations in the ventral visual system that are transform invariant by using information that over short periods of time, the object is likely to be the same, but seen in different transforms (Földiák, 1991; Rolls, 1992; Rolls, 2012b; Wallis and Rolls, 1997), but is applied to computations in several stages of the dorsal visual system here. The concept for the dorsal visual system is that the same location in allocentric space may be viewed for short periods in which the viewing may involve a number of different retinal positions, eye positions, head directions, and even places if one is walking, so that slow learning taking advantage of these statistics of the environment may help one to build representations of for example a location out there that are invariant with respect to retinal and eye position, head direction, and the place where one is located. It is proposed in this paper that this memory trace learning principle could be applied at a number of stages of processing involving different coordinate transforms in the dorsal visual system. Some progress has been made in applying this to one stage of processing, that involved in forming head-centered representations (Navarro et al., 2018).

Gain modulation to produce coordinate transforms is a well-established principle of operation of neuronal systems in the dorsal visual system (Salinas and Abbott, 1995; Salinas and Abbott, 1996; Salinas and Abbott, 2001; Salinas and Sejnowski, 2001). The term gain field describes the finding that the response of a neuron in parietal areas 7a, LIP and VIP to a visual stimulus at a given position on the retina (the neuron's receptive field) can be modulated (decreased or increased) by a modulating factor, eye position (the angle of the eye in the head) (Andersen, 1989; Andersen et al., 1985; Andersen and Mountcastle, 1983; Duhamel et al., 1997). Each neuron thus responds best to a combination of retinal and eye position. The gain modulation by eye position occurs in a spatially systematic and nonlinear way such that the output of the population of neurons encodes the position of the stimulus relative to the head, by taking into account both retinal position and eye position (Salinas and Abbott, 2001; Salinas and Sejnowski, 2001). This gain modulation can be thought of as shifting the retinal receptive field of the population of neurons so that they represent direction relative to the head, which is a spatial coordinate transform (see illustration in Fig. 3a).

A problem with the gain modulation mechanism in practice is that it may not be perfect at each stage (Graf and Andersen, 2014), and when successive stages involving other coordinate transforms follow, the imperfections at each stage combine to make a system that operates very imperfectly, as is shown by the simulations described later. It is proposed here that a temporal trace synaptic learning mechanism can help by using the statistics of the natural world across time to help with the learning. In the example that we have, a visual stimulus might be steady at a given position relative to the head for several seconds during which many eye movements would occur. The eye movements would result in different combinations of eye position and retinal position occurring in those few seconds. If the active neurons maintained a short-term memory trace of recent neuronal or synaptic activity for a short period, of even a few hundred ms, the neurons could then learn about what was constant over short periods (such as the position of the visual stimulus relative to the head). The trace learning mechanism itself is very simple and biologically plausible, for it can be included in a competitive network, a standard network in cortical systems, just by utilizing for example the long time constant of NMDA receptors, or the continuing firing of cortical neurons for 100 ms or more that is characteristic of cortical networks with recurrent connections to form attractor networks (Földiák, 1991; Franzius et al., 2007; Rolls, 1992; Rolls, 2012b; Rolls, 2016a; Wallis and Rolls, 1997; Wallis et al., 1993; Wiskott and Sejnowski, 2002; Wyss et al., 2006). Exactly these cortical processes provide a theory and model for transform-invariant object representations in the ventral visual system (Eguchi et al., 2016; Rolls, 1992; Rolls, 2012b; Rolls, 2016a; Rolls and Mills, 2018; Wallis and Rolls, 1997; Zhao et al., 2019), and are now proposed here to play an important role in several stages of the dorsal visual system, in relation to learning spatial coordinate transforms. The neural mechanisms will be specified formally and mathematically in the Methods section.

In summary, it is proposed here that spatial coordinate transforms in the dorsal visual system, and potentially in other neural systems, are learned by a combination of gain modulation, a short term memory trace rule learning, and competitive learning to select neurons at each stage of the hierarchy. The processes may take place during post-natal development.

The theory for a set of spatial coordinate transforms performed in the dorsal visual system

In line with empirical evidence on some of the spatial coordinate frameworks present in the primate dorsal visual system and related parietal cortex areas (Andersen and Cui, 2009; Bremmer et al., 2002; Bremner and Andersen, 2014; Byrne et al., 2007; Duhamel et al., 1997; Epstein, 2008; Galletti and Fattori, 2018; Snyder et al., 1998; Vann et al., 2009; Vedder et al., 2017; Whitlock, 2017), and the primate hippocampus (Feigenbaum and Rolls, 1991; Georges-François et al., 1999;

Robertson et al., 1998; Robertson et al., 1999; Rolls et al., 1997a; Rolls et al., 1998; Rolls and Wirth, 2018; Rolls and Xiang, 2005; Rolls and Xiang, 2006; Rolls et al., 2005; Wirth et al., 2017), the following set of spatial coordinate transforms are considered here, and are specifically investigated in the 3-layer model of successive coordinate transforms described below and illustrated in Fig. 2. Each layer in the model corresponds to a different cortical processing area. The principles may, it is postulated, apply to other spatial coordinate transforms present in these and other cortical areas.

Layer 1: Retinal position is the input to this layer or cortical area, where it is gain modulated by eye position to produce position with respect to the head. The coordinate framework thus becomes head-centred, as illustrated in Fig. 3a. Competitive learning with a temporal trace learning rule to select the Layer 1 neurons with good responses for each position with respect to the head occurs, learning over all combinations of retinal position and eye position that correspond to a given position in head-based coordinates. This enables neurons to respond to a given position in head-centred space over many combinations of retinal and eye position that correspond to that position in head-centred space. Neurons that respond in head centered coordinates are found in macaque areas VIP and LIP (Andersen, 1989; Andersen et al., 1985; Andersen and Mountcastle, 1983; Duhamel et al., 1997).

Layer 2: Layer 2 receives the head-centred representation from Layer 1, and utilizes gain modulation by head direction to produce a representation of direction in space that is independent of head direction. This can be thought of as bearing direction to a landmark, which is allocentric, as described in the Introduction and as illustrated in Fig. 4a. Competitive learning occurs with a temporal trace learning rule to select the Layer 2 neurons with good responses for this bearing direction in space, learning over all combinations of retinal position, eye position, and head direction that correspond to a given bearing direction from the primate at which a stimulus is present. A representation of this type may be present in primate parietal area 7a (Snyder et al., 1998) and the posterior cingulate cortex (Dean and Platt, 2006), as considered in the Discussion. (Cells in the rat medial entorhinal cortex that code the allocentric bearing to an object (Wang et al., 2018), and to a goal by CA1 cells of bats (Sarel et al., 2017), may be analogous.) Bearing cells need to be selective for different landmarks, if they are to be used to generate spatial view cells with specificity for particular parts of a scene, such as a landmark in the scene. Selectivity of bearing cells for landmarks also allows them to be used to calculate place by triangulation. This Layer 2 bearing direction representation would also be useful for navigation, as well as part of the input to build further representations in higher layers. Bearing cells might encode distance to the landmark as well as bearing, and if so, the distance from as well as the bearing to the landmark would be gain-modulated out in Layer 3.

   This Layer 2 representation would also be useful for reaching to a location in space independently of head direction, and for that matter eye position and retinal position. The location out there in space defined in this way would still depend on the place at which the primate is located, and the location is not yet an allocentric representation of a particular location out there in space. To make an actual reach movement, a transform into body-centred space would be needed, as the arms are anchored to the body. Visual neurons that respond in body-centered coordinates have been recorded in area LIP (Snyder et al., 1998). Such body-centred neurons useful for reaching could be produced by the mechanisms described for Layer 2 of the present model, in which the head-centred representation is gain modulated by body rotation.

Layer 3: The bearing direction representation of a stimulus or landmark from Layer 2 is then gain modulated by place in Layer 3 to produce a view representation of a location in space that is independent of the place of the primate in space. (In this paper, 'location in space' refers to the allocentric location being viewed in a spatial scene; and place refers to the allocentric place where the viewer is.) The location out there in space would therefore be in allocentric, that is, world-based, coordinates, as illustrated in Fig. 5a. This is the same coordinate frame as spatial view cells in the primate hippocampus (Georges-François et al., 1999; Kesner and Rolls, 2015; Rolls and Wirth, 2018; Rolls and Xiang, 2006), and enables an interface in the same coordinate frame between the dorsal visual system and the primate hippocampal system. Competitive learning with a temporal trace learning rule to select the Layer 3 neurons with good responses for allocentric representations of spatial view is also implemented in Layer 3, learning over all combinations of retinal position, eye position, head direction, and place that correspond to a given view of a location in allocentric space. Connections reach parietal cortex area 7a from the dorsal visual stream areas. Area 7a connects on to areas such as the retrosplenial cortex and posterior cingulate cortex (Kobayashi and Amaral, 2003; Kobayashi and Amaral, 2007), which provide access to the hippocampal system (Fig.

1). This route potentially enables an allocentric visual representation of space derived from the dorsal visual system to be interfaced to the hippocampal allocentric spatial view system, which provides information about the location of objects and rewards and connected locations in the world (Kesner and Rolls, 2015; Rolls, 2016a; Rolls, 2018; Rolls and Wirth, 2018). Effectively, Layer 3 computes bearing invariance to encode a location out there in space that is bearing invariant, which is important for remembering the location where objects or rewards have been seen, and generalization across different bearings.

These transforms, and how they are produced, will be evident when the operation of the model of these processes is described in the Results section.

3. Methods

The computational model for successive spatial coordinate transforms (VisNetCT) is very similar to the VisNet model for the ventral visual system (Rolls, 2012b; Rolls, 2016a; Rolls and Milward, 2000), except that gain modulation is added at each layer or stage or cortical area of processing, and except that the convergence from stage to stage is more limited so that topology is maintained.

The network architecture is illustrated in Fig. 2, and is a series of competitive networks, organized in hierarchical feedforward layers. There is mutual inhibition of neurons within each layer to implement competition and maintain a sparse representation. Each layer operates using competitive learning (Rolls, 1989a; Rolls, 2016a; Rumelhart and Zipser, 1985; von der Malsburg, 1973) which is helped by the diluted connectivity present in VisNetCT (Rolls, 2016b). A modified associative (Hebb-like) synaptic learning rule incorporating a temporal trace of each neuron's previous activity is used at each stage. The details of the rate model follow, with further details about the general model VisNet elsewhere (Rolls, 2012b; Rolls and Milward, 2000; Wallis and Rolls, 1997).

The temporal trace learning rule

The trace learning rule utilizes the spatio-temporal constraints that are often present for stimuli in the environment. An example has been provided above in which the position of a stimulus may be constant with respect to head position during which a number of different combinations of retinal and eye positions may be associated with a single direction with respect to the head. The problem is to learn transform invariance for the different retinal and eye position transforms, to compute a representation that is invariant with respect to these combinations or transforms. This is the analogy with VisNet used in the ventral visual system, in which the transforms might be the location and view of an object, and we wish to learn a representation of the object that is invariant with respect to these transforms (Rolls, 2012b; Rolls, 2016a).

The trace update rule is

$$\delta w_j = \alpha \bar{y}^\tau x_j \tag{1}$$

where

$$\bar{y}^\tau = (1-\eta)y^\tau + \eta \bar{y}^{\tau-1} \tag{2}$$

and

x_j is the jth input to the neuron;

y is the output from the neuron;

$\bar{y}^\tau$: is the Trace value of the output of the neuron at time step τ ;

α is the learning rate;

w_j is the synaptic weight between j^{th} input and the neuron;

η is the trace update proportion, with 0 meaning no trace, just associative learning. The optimal value varies with the presentation sequence length, and is typically 0.8.

At the start of a series of investigations of different forms of the trace learning rule, we demonstrated (Rolls and Milward, 2000) that VisNet's performance could be greatly enhanced with a modified Hebbian trace learning rule (equation 3) that incorporated a trace of activity from the preceding time steps, with no contribution from the activity being produced by the stimulus at the current time step. This rule took the form

$$\delta w_j = \alpha \overline{y}^{\tau-1} x_j^{\tau} \tag{3}.$$

The trace shown in equation (3) is in the postsynaptic term. The crucial difference from the earlier rule (see equation (1)) was that the trace should be calculated up to only the preceding timestep, with no contribution to the trace from the firing on the current trial to the current stimulus. This has the effect of updating the synaptic weights based on the preceding activity of the neuron, which is likely given the spatio-temporal statistics of the visual world to be from previous transforms of the same object / stimulus (Rolls and Milward, 2000; Rolls and Stringer, 2001). This is biologically plausible, as considered in more detail elsewhere (Rolls, 2012b; Rolls, 2016a), and this version of the trace rule was used in this investigation.

To bound the growth of each neuron's synaptic weight vector, $\mathbf{w}_i$ for the i th neuron, its length is explicitly normalized (a method similar to that employed by von der Malsburg (1973) and Rumelhart and Zipser (1985) which is commonly used in competitive networks (Rolls, 2016a)). An alternative, more biologically relevant implementation, using a local weight bounding operation which utilizes a form of heterosynaptic long-term depression (Rolls, 2016a), has in part been explored using a version of the Oja (1982) rule (see Wallis and Rolls (1997)).

The synaptic learning rate α, the amount that the synaptic weight altered when presented with a stimulus, was adjusted to a value (typically 0.05) that ensured that the synaptic weights converged to steady values after a number of training trials.

The network implemented in VisNetCT

The network itself is designed as a series of hierarchical, convergent, competitive networks. The network implemented consists of a series of three layers, constructed such that there is some convergence from layer to layer, as illustrated in Fig. 2. The forward connections to a neuron in one layer are derived from a topologically related and confined region of the preceding layer. The choice of whether a connection between neurons in adjacent layers exists or not is based upon a Gaussian distribution of connection probabilities that roll off radially from the focal point of connections for each neuron. (A minor extra constraint precludes the repeated connection of any pair of cells.) In particular, the forward connections to a neuron in one layer come from a small region of the preceding layer defined by the radius which will contain approximately 67% of the connections from the preceding layer. This radius was set to 2 for each layer of 32x32 neurons per layer, and each neuron received 100 synaptic connections from the neurons in the preceding layer. This resulted in the maintenance of some topology through the different layers of VisNetCT, which is different from VisNet, in which the aim is to produce neurons in the final layer with full translation (shift) invariance as well as other invariances (Rolls, 2012b). In the rate model the activation of a neuron in a layer is calculated as a weighted sum of the inputs it receives from the preceding layer multiplied by the corresponding synaptic weights, i.e. as a dot or inner product, with the details of the implementation of VisNet described by Rolls and Milward (2000). The activations were converted into firing rates using a threshold-linear activation function and the method described next to produce a given sparseness of the representation in a layer.

Competition and mutual inhibition in VisNetCT

In order to act as a competitive network some form of mutual inhibition is required within each layer, to ensure that only a proportion of neurons is active for any one stimulus. After the activation of the neurons in a layer had been calculated by the dot product of the synaptic weights of a neuron and the firing rates of the neurons in the preceding layer to which it was connected by the synaptic weights, the activations were converted into firing rates in VisNetCT using a threshold linear activation function with the threshold set for each firing rate update of a layer so that the sparseness of the firing became a fixed value specified by a sparseness parameter a that was typically 0.008, where sparseness is defined as

$$a = \frac{(\sum_i y_i / n)^2}{\sum_i y_i^2 / n} \tag{4}$$

where n is the number of neurons in the layer, and y_i is the firing rate of the i th neuron in a layer. This sparseness measure is one useful in the quantitative analysis of the capacity of neuronal networks (Rolls, 2016a; Rolls and Treves, 1990; Treves, 1991; Treves and Rolls, 1991), and in neurophysiological measures of neuronal representations in the brain (Franco et al., 2007; Rolls,

2016a; Rolls and Tovee, 1995). The typical sparseness value of 0.008 resulted in approximately 8 neurons having high firing rates in a layer, given that there were 1024 neurons in each layer. In VisNetCT for simplicity the inhibition within a layer implemented in this way was global, across all neurons.

Gain modulation at each stage, and the input stimuli in VisNetCT

The gain modulation was performed in the general way described elsewhere (Salinas and Abbott, 2001; Salinas and Sejnowski, 2001). For simplicity, the gain modulation was implemented by a convolution of firing in a layer with the gain modulation signal. For example, the stimulus was presented on Layer 1 shown in Fig. 1, which might correspond to an area such as LIP or earlier in the dorsal visual system. Each stimulus was a set of approximately 4 pixels with high firing in retinal coordinates on the input layer (produced by convolving a single pixel of activity with a Gaussian filter using a Gaussian kernel with radius 1 pixel). To produce the gain modulation by eye position, the firing in Layer 1 was convolved with the eye position, and thereby shifted across the population of neurons so that the peak firing in Layer 1 then represented the retinal input now gain modulated by eye position to represent position relative to the head. A series of inputs corresponding to one location in head centred space but produced by different combinations of retinal position for the stimuli and eye position were presented successively, so that the trace rule could encourage neurons in Layer 1 to learn the different retinal and eye position combinations that produced that position in head direction space. Then all the combinations that corresponded to the next position in head-based space were presented close together in time, etc. The trace was reset between each new coordinate to be learned, and trace rule learning was not allowed to occur until stimuli had been presented 4 times, to enable the trace firing to settle down. The whole set of head-based representations was trained for typically 5-50 such epochs in Layer 1.

Similarly, for Layer 2, the gain modulator was the head direction, and this was used to convolve the Layer 2 firing into approximately bearing-related coordinates (i.e. for a direction in space that was independent of head direction). Trace rule learning was used every time a new stimulus was presented, which consisted of all combinations of retinal and eye position, and head direction, that corresponded to a single bearing direction. That was repeated for the other bearing directions. And then that single epoch of training was repeated typically for 5-50 epochs.

The output from Layer 2 was used as the input to Layer 3, for which the gain modulator was the place at which the agent was located, and this was used to convolve the Layer 3 firing into approximately spatial view coordinates (i.e. for a position in allocentric space that was independent of the place where the agent was located). Layer 3 also performed trace rule learning over several epochs in which in each epoch and for each spatial view all combinations of now retinal position, eye position, head direction, and place that corresponded to a single location in allocentric space "out there", the spatial view, were presented close together in time. (In practice, with VisNet each training epoch need not contain every possible combination to be learned, as the continuity can be built up over a number of training epochs in each of which there is some continuity (Rolls, 2012b).)

The computer algorithm that implemented each stage of processing operated as follows to combine the gain modulation and competitive learning using a trace rule, using Layer 2 above as an example. An input stimulus was applied, and after the firing of Layer 1 (in head-based coordinates) was computed, gain modulation by convolution with head direction was applied to this representation in Layer 1. That Layer 1 firing then activated Layer 2 neurons through the synaptic weights from Layer 1 to Layer 2, competition was implemented by mutual inhibition of the Layer 2 neurons, and the resulting firing was combined with the firing from the preceding few trials, with the resulting short-term memory trace of recent neuronal firing used to update the synaptic strengths between Layer 1 and Layer 2 as described in Equations 1-3. As described in the Discussion, the time scale over which the trace rule needs to operate for these transforms in the real world may be shorter for early stages of the system (in which eye movements have a time scale of fractions of a second) compared to later stages (in which locomotion could take place over seconds or more). For longer time scales, the trace rule could be helped by the longer times scales of neuronal activity provided for by attractor networks implemented by the recurrent collateral connections between nearby cortical neurons (Rolls, 2012b; Rolls, 2016a).

Parameters used in VisNetCT

The rate model used three layers each composed of 32 x 32 neurons, each with 100 connections to neurons in the preceding layer selected using a Gaussian probability distribution with a radius of 2. The input stimulus was a single pixel convolved with a 2D Gaussian with standard

deviation of 1, producing a 2D stimulus that was several pixels in diameter. It could be placed at one of three positions of the retina, at X=-5, 0, and +5 with Y=0 throughout. Three eye positions, three head directions, and three places were used as gain modulators in layers 1, 2, and 3 respectively of the architecture as described. The learning rate parameter α for the synaptic update was 0.05, the trace learning rule parameter η was 0.8, and 5-50 training epochs were typically run, although the net typically reached an asymptote of performance much sooner. The sparseness *a* of the representation in a layer was 0.008. Although for simplicity the model was trained in only the horizontal (X) plane, the model is 2D to provide for both the X and Y dimensions, to allow the connectivity of a single neuron to come from a 2D Gaussian region of the preceding layer with realistic diluted connectivity, and because the stimulus was 2D because it was a single pixel convolved with a 2D Gaussian with standard deviation of 1.

Measures for network performance
Information theory measures
  The performance of VisNetCT was measured with Shannon information theory (Shannon, 1948) using methods that are identical to those used to quantify the specificity and selectiveness of the representations provided by neurons in the brain (Rolls, 2012b; Rolls, 2016a; Rolls and Milward, 2000; Rolls and Treves, 2011). A single cell information measure indicated how much information was conveyed by the firing rates of a single neuron about the most effective stimulus. A multiple cell information measure indicated how much information about every stimulus was conveyed by the firing rates of small populations of neurons, and was used to ensure that all stimuli had some neurons conveying information about them.
  A neuron can be said to have learnt an invariant representation if it discriminates one set of inputs from another set, across all transforms. For example, a neuron's response in Layer 3 of VisNetCT is invariant for spatial view if it responds to primarily one spatial view and not other spatial views, independently of all the combinations of retinal position for the stimulus, eye position, head direction, and place where the agent is located. The responses of each neuron in for example Layer 3 were found for each of the spatial views, and the amount of information the single cell had about spatial view was measured. In the present context, the measure provides a metric for measuring how many spatial views could be represented. The algorithms and their use are described in detail elsewhere (Rolls and Milward, 2000; Rolls and Stringer, 2006; Rolls et al., 1997b; Rolls et al., 1997c).

Correlation matrix between the firing of neurons in a layer
  After assessing the results with these information theoretic methods, it was also useful to show how the different layers of the network categorised the stimuli by computing a correlation matrix based on the firing rates of all the neurons in a layer for every combination of the stimuli and modulators that were involved. If this matrix was set up with the combinations corresponding to, for example, the different combinations for one spatial view, then the combinations for the next spatial view, then this could clearly show whether spatial view was encoded within a layer, as illustrated in Figs. 3-5. It was also useful to show how single neurons in a layer categorised the stimuli by showing their firing rates as functions of for example spatial view and in other coordinate frames, as shown in Fig. 6.

Results
  The concepts in the approach described here are illustrated by simulations of the architecture shown in Fig. 1. The operation of the networks illustrated in Fig. 1, and what is computed at each stage and how it is computed, is described by starting with a description of the modelling for Layer 1, and working up through the system to Layer 3. The parameters that were used for the simulations were as shown in Table 1 except where otherwise stated.

Layer 1
  For Layer 1, the training involved presenting the stimulus at all combinations of 3 retinal positions with 3 eye positions as gain modulators. This resulted in a conversion to head-centered coordinates as illustrated schematically in Fig. 3a. The details of how each combination that corresponded to a given head-centred position were presented close together in time to enable the trace rule to contribute to the learning are described in detail in the rest of this paragraph. The results obtained are described in the paragraphs that follow. For Layer 1 the input stimulus to Layer 1 was shown in 3 retinal positions in the 32x32 spatial grid at X = -5, 0 and 5 (where 0 corresponds to the

centre, and X signifies the horizontal plane). The eye positions used were with the eyes at three deviations in the horizontal plane, corresponding to the fovea looking at X=-5, X=0, and +5. Training was performed with the stimulus placed first at one position in head-centred space, with all the combinations of retinal position and eye position that corresponded to that position in head-based space presented close together in the sequence to enable the trace learning rule to encourage neurons in Layer 1 to learn that all these combinations were about the same location in head centred space. Then all combinations of retinal and eye positions corresponding to another position in head centred space were presented, to enable the trace rule learning to allocate neurons to all the combinations of retinal and eye position that corresponded to that position in head-centred space to be learned. There were in all 5 positions in head-related space that were defined by these combinations of retinal and eye position. (They are X = -10, -5, 0, +5, and +10 in head-centred space.) The computer algorithm performed the gain modulation before the input reached Layer 1, and Layer 1 implemented the trace rule learning to learn to allocate neurons to each of the five positions in head-centred space. The computer algorithm also ensured that the head positions were chosen in permuted sequences. (It should be remembered that the stimulus was a Gaussian blob that covered several pixels on the retina, and that the connectivity was diluted and probabilistic as in the cortex, so that the neurons in the receiving layer had to perform a non-trivial computation to learn to allocate different neurons to each position in the space being represented in each layer.)

The results of this training for Layer 1 are illustrated with the correlation matrix shown in Fig. 3b. This represents the correlation of the firing between all the neurons in Layer 1 for each of the five head centred positions at which a stimulus was presented. This shows that the 5 different head-centred positions for a stimulus each produced different firing of Layer 1 neurons, even though each head-centred position was produced by a number of combinations of retinal stimulus and eye position, as explained in the legend to Fig. 3.

The simulation of Layer 1 thus shows that the competitive learning in Layer 1 using a short-term memory trace in combination with gain modulation by eye position can learn to allocate different neurons to respond to each head-centred position in space. The task becomes more demanding, and the trace learning rule makes a more important contribution, as we progress through the layers.

Layer 2

For Layer 2, the five head-centred outputs of Layer 1 were gain-modulated by head direction to produce firing that was related to the bearing to a landmark L as illustrated in Fig. 4a, using the trace learning rule in the competitive network of Layer 2. There were three head-direction modulators, corresponding to the head directed towards X=-5, X=0, and X=+5. This produced seven possible bearing directions each to a different landmark L, from the single place where the viewer was located. The correlation matrix shown in Fig. 4b represents the correlation of the firing between all the neurons in Layer 2 for every one of the seven bearing directions to a landmark relative to the agent at which a stimulus was presented, after training Layers 1-2. This shows that the 7 different bearing directions for a stimulus each produced different firing of Layer 2 neurons, even though each bearing direction was produced by a number of combinations of retinal stimulus, eye position, and head direction.

Layer 3

For Layer 3, the seven bearing direction outputs of Layer 2 to different locations L were gain-modulated by place to produce an allocentric view-based representation using the trace learning rule in the competitive network of Layer 3. This might be produced by walking from one place to another while watching the same location in a scene. Fig. 5a shows a schematic for a single location L in the scene. Each location was encoded by spatial view cells that encoded the location in the scene, independently of the place where the viewer was located. There were three place modulators, corresponding to the places in the horizontal plane of X=-5, X=0, and X=+5. This produced representations of 9 allocentric, world-based, views each of a different location in the scene, corresponding to -20 and +20 (1 combination each), -15 and +15 (4 combinations each; -10 and +10 (10 combinations each); -5 and + 5 (16 combinations each); and 0 (19 combinations). Each of the spatial view representations learned was of a different location L in the scene. Each spatial view representation was independent of the place where the viewer was located. A neuron that responds allocentrically to a single location in the scene is termed a spatial view cell (Georges-François et al., 1999).

Fig. 5b shows that after training Layers 1-3, the firing of neurons in Layer 3 represented each of the spatial locations in the scene separately and independently. The correlation matrix represents the correlation of the firing between all the neurons in Layer 3 for every spatial location presented. The stimulus order for this diagram is provided in the legend to Fig. 5, but results in spatial view 1 (i.e. for landmark 1) being the block at the top left, and spatial view 7 (for landmark 7) being the block in the bottom right. The first spatial location (for X= -15) was represented by 4 combinations of the retinal position and modulators of eye position, head direction, and place), and formed one block of correlated representations in the top left of Fig. 5. Fig. 5 thus shows that the firing of Layer 3 represented each spatial location in the scene almost orthogonally to the other spatial locations, which is good performance. (Firing for spatial location coordinates of -20 and +20 were not included in Fig. 5, because there were too few combinations of stimuli and gain modulators at these extremes.)

The information theoretic analysis for Layer 3 showed that the average information about spatial location in the scene for the 5 neurons most selective for each of the 7 spatial locations was 2.42 bits (where 2.81 bits is the maximal single cell information value calculated as $\log_2$ of the number of stimuli). As a control, the value without training was much lower, 1.56 bits. This is not zero, because by chance some neurons with the initially random synaptic connections will have connectivity that allows some effect of one layer on the next, and if those neurons can be found during the training, they will respond to some of the stimuli better than other stimuli.

The responses of a Layer 3 single neuron selective for a spatial location in the scene (i.e. a spatial view cell) after training are shown in Fig. 6. The top left panel shows that the average rate of the neuron to every combination of retinal position, eye position, head direction, and place that corresponded to spatial location 5 was 16, and that the rate was zero for all other spatial views. The other panels show that the same spatial view neuron did not discriminate between the place where the viewer was, head direction, or eye position (which is also the case for spatial view cells in the primate hippocampus (Georges-François et al., 1999)). Fig. 6 also shows that the spatial view neuron did not discriminate between retinal positions. The bottom left panel of Fig 6 shows that the same Layer 3 spatial view neuron did not have firing that was specific to a given bearing direction to a landmark (which, as was shown, was encoded by Layer 2 neurons), and the bottom right panel shows that the neuron was not specific to position relative to the head (which, as was shown, was encoded by Layer 1 neurons). (The different firing shown in the bottom two panels just reflected how frequently these egocentric and head centred coordinates occurred when spatial view 5 was being tested, and that the neuron was tuned to spatial view 5 as shown in the top left panel.)

To provide evidence on the utility of the trace rule learning used in combination with gain modulation described here, the whole simulation was rerun without the trace rule, and using instead the conventional Hebbian associative rule with no temporal trace learning for the competitive networks (Rolls, 2016a). The simulation was identical to that used to produce the results in Fig. 5 apart from that change to the learning rule. The results in Fig. 7 show that Layer 3 of the network categorised the stimuli into different spatial views much less well than with the trace rule (which is what is shown in Fig. 5). That is, for each of the 7 spatial locations, the different combinations of retinal position, eye position, head direction, and place produced rather different firing of the Layer 3 neurons, showing that without the trace rule, and only with gain modulation, the coordinate transform network worked much less well. This analysis was confirmed by the result that the average single cell information of the 5 best neurons for each spatial location was now only 1.6 bits.

Discussion

In the architecture for the dorsal visual system described here, Layer 1 uses gain modulation of a retinal signal by eye position, and trace learning, to compute representations that are invariant with respect to retinal and eye position, producing head-centered representations. Layer 2 uses gain modulation by head direction and trace learning to compute representations that are invariant with respect to head direction, producing bearing representations to a given stimulus or landmark. Layer 3 uses gain modulation by place and trace learning to compute representations that are invariant with respect to bearing and place (see Fig. 5a), which is important for remembering the location 'out there' where objects or rewards have been seen, and for generalization across different bearings to a location in a scene, and across different places from which the scene is viewed. This is a particular instantiation of a general architecture for coordinate transforms in the brain, with similar

mechanisms of gain modulation and trace rule learning involved at each stage. Many interesting principles are involved, and are considered next.

First, a new approach to coordinate transform learning in the brain has been proposed, in which the more traditional gain modulation is complemented by temporal trace rule competitive neuronal network learning to learn coordinate transforms in many regions in the dorsal visual system. It is shown that the new approach works much more satisfactorily than gain modulation alone (compare Figs. 5 and 7). Further, this approach provides a mechanism for individual neurons to represent many of the input combinations of for example eye and retinal position that correspond to a single direction in head-based coordinates. This understanding may have application to many brain areas where coordinate transforms are learned.

Second, a set of coordinate transforms has been proposed for the dorsal visual system / parietal areas that enables an allocentric representation to be formed, of the location in a scene. The input stimulus is merely a stimulus at a given position in retinal space, and the gain modulation signals needed are eye position, head direction, and place, all of which are signals present in the primate brain (Chen et al., 2018; Hazama and Tamura, 2019; Robertson et al., 1999; Rolls and O'Mara, 1995; Rolls and Wirth, 2018; Whitlock, 2017; Wirth et al., 2017). Part of the interest here is that the allocentric spatial representations of a location 'out there' in space produced in the model described here are in the same coordinate framework as spatial view cells recorded in the primate hippocampus and parahippocampal cortex (Georges-François et al., 1999; Robertson et al., 1998; Rolls et al., 1997a; Rolls et al., 1998; Rolls and Wirth, 2018). It is of course the case that two of the modulators are allocentric, place and head direction.

Third, the particular coordinate transforms considered here involved gain modulation at successive stages of processing in the dorsal visual stream using as gain modulators eye position, head direction, and place. The theory and model described here is intended to be a conceptual approach to how a series of coordinate transforms could be implemented by biologically plausible neuronal networks in the dorsal visual system and parietal cortex, with the actual gain modulators perhaps being different to produce other coordinate transforms. In particular, in the third stage shown in Fig. 2, the gain modulator is place to produce a view representation, but there could be other gain modulators producing different coordinate transforms in parts of the system. For example, another gain modulator might be body rotation, which does modulate the responses of some LIP neurons (Snyder et al., 1998), and that could be useful in Layer 2, leading for example to a representation in a body-related coordinate frame, which would be suitable for directing arm movements to positions in space, given that the arm muscles are anchored to the body.

Fourth, we should consider how some of the representations described could be produced by neuronal network operations. We have argued that spatial view cells found in the hippocampus and parahippocampal cortex could be produced by neurons learning to respond to a combination of the visual features present over a small angle of perhaps 10-30 degrees close to the fovea, and we have produced a formal model of this (de Araujo et al., 2001). Such a representation would include detailed information about the stimuli or features present, such as a house in a scene to the left of a church. Each such spatial view would be linked to the next spatial adjacent and partially overlapping spatial view in a continuous attractor network in ways described in detail elsewhere (Rolls, 2016a; Stringer et al., 2005). For a given environment this can be described as a chart, and the capacity of a neural network for storing many such charts has been calculated (Battaglia and Treves, 1998), and is high because the whole charts of different environments are relatively uncorrelated. The continuous attractor / chart / schema that represents the structure of a maze and the body turns needed to reach a goal could remain the same if the wall cues were changed, and the chart could be oriented using the goal as a reference after the change of wall cues, with the new wall cues / landmarks potentially added to the existing continuous attractor / chart / schema representing the maze (Baraduc et al., 2019). The place where the animal is located could be incorporated in such continuous attractor networks, as well as the room cues, and body turns needed as the maze is navigated (Rolls and Wirth, 2018). Consistent with this, place cells, that is, neurons that fire when the macaque is at a given place, are found in the primate hippocampus (Hazama and Tamura, 2019; Rolls and O'Mara, 1995). These place cells could be produced by neurons learning to respond to a combination of the environmental landmarks present (such as may be implemented by ventral visual stream scene cells (Kornblith et al., 2013; Nasr et al., 2011)) together with their bearing, and we have produced a formal model of this (de Araujo et al., 2001) and referred to this previously (Rolls and Wirth, 2018). Primate hippocampal neurons that respond to landmarks but are place-dependent in a virtual reality task (Wirth et al., 2017) have some similarity to the 'bearing to a location' cells described here for Layer 2. In addition, the rodent CA1 'landmark vector cells' (Deshmukh and Knierim, 2013) which respond

at a particular bearing and distance from an object (sometimes called a 'landmark') within an enclosure may be analogous to the allocentric 'bearing to a location' cells described here, though in the primate the bearing cells are likely to be anchored to consistent landmarks that define the space and can be fixated exactly with the primate visual system, and less to objects placed into a space. One way in which primate (including human) place representations could be formed is by the system described by de Araujo, Rolls and Stringer (2001) in which place cells can be formed by using a combination of landmarks and their bearings in a process that is essentially triangulation and that is implemented by competitive learning. A complementary mechanism could be a system that can produce place representations from idiothetic grid cell representations using competitive learning in the dentate gyrus / hippocampal system (Rolls et al., 2006).

Fifth, the new concepts introduced here extend the use of trace learning from the ventral visual system for transform-invariant visual object recognition (Földiák, 1991; Franzius et al., 2007; Rolls, 1992; Rolls, 2012b; Rolls, 2016a; Wallis and Rolls, 1997; Wallis et al., 1993; Wiskott and Sejnowski, 2002; Wyss et al., 2006) to multiple layers of the dorsal visual system. This helps to emphasize the powerful nature of the training signals that can be derived from the temporo-spatial statistics of what reaches the brain from the world, which as shown here can be useful for helping to learn coordinate transforms. There may be other applications of temporal trace learning in the brain (Franzius et al., 2007; Zhao et al., 2019), as these provide a type of teaching signal for neurons. The trace rule learning in combination with gain modulation provides especial help in the following two ways. First, the trace rule applied separately to each neuron encourages single neurons to learn the different combinations of their inputs provided by different combinations of the signal to be modulated and the gain modulator signal. Thus for example a neuron in Layer 1 can respond to several different combinations of retinal and eye position that correspond to the same location in head-centred space. This makes the readout of the information much more reliable, because the neurons that read the output will then respond to any of the many different transforms that have been learned, and thus generalize well across the different combinations (Rolls, 2016a). Second, the competitive network that is implemented as part of the process helps to sharpen up the representations, so that each neuron tends to respond to a different part of the transformed space, for example to a different direction in head-based space in Layer 1. This is a decorrelation or orthogonalization process, which is sometimes referred to as pattern separation, and which helps to make the transmission and storage of information have a high capacity (Marr, 1971; McNaughton and Morris, 1987; Rolls, 1987; Rolls, 1989a; Rolls, 2016b; Treves and Rolls, 1992). The whole philosophy here is quite different to that involved in deep learning using backpropagation of error (LeCun et al., 2015). Instead, here we have a system that self-organises without explicit teachers for every neuron, by using instead the temporo-spatial continuity that is present in the world, and bottom-up feed-forward self-organizing learning.

For the trace rule learning to be useful, the statistics of the world do have to be appropriate, with, for example, for head-centred learning an object or stimulus to be present at a given head-centred position while eye movements are made, so that different combinations of retinal position and eye position occur close together in time. During development, there may be separate periods in which different types of plasticity occur, such that some coordinate transforms are set up before others. Different time constants of learning in the system at different stages may also enable the brain to extract relevant statistics from the world for the learning of different coordinate transforms. For example, eye position changes have a relatively fast time course with often 2 saccades per second (Roberts et al., 2013), whereas head direction changes may be on a slower time scale of 1 to several seconds, and movement to different places may be on a slower time scale of seconds to hundreds of seconds or more. There is plenty of opportunity here for experimental investigations to explore these issues. However, this proposal does seem feasible, in that natural visual stimuli can be sufficiently rich for useful information to be extracted to facilitate learning of different representations depending on the stage in the hierarchy of visual processing areas (Wyss et al., 2006).

In the present approach, the gain modulation is important, but so is the competitive learning from stage to stage of the hierarchy (see Fig. 1), for that is an important mechanism for allocating neurons to represent the input information efficiently by different neurons (Rolls, 2016a). In this context, the addition of a short-term memory trace to the competitive learning is shown to have advantages, because it enables neurons to learn what is constant over short time periods. For example, in Layer 1, neurons can learn to respond to a stimulus at the same head-centred location for several seconds, by learning to respond to several combinations of retinal and eye position that may occur in that time.

It is of interest to compare the learning of invariant transforms in the ventral and dorsal visual system, and how slow learning implemented by the trace learning rule may contribute to both. In the primate ventral visual system, the trace rule may allow different transforms of an object to be associated together, including position, size, and even view which can be completely different for the same object seen from different views (DiCarlo et al., 2012; Rolls, 1992; Rolls, 2012b; Rolls, 2016a; Wallis and Rolls, 1997; Wiskott and Sejnowski, 2002). Effectively, the representation of an object is made position, size, view, lighting etc invariant, so that when the object is associated with reward in the orbitofrontal cortex or enters the episodic memory system in the hippocampus, generalization over all these transforms occurs when the object is encountered again but in a different transform (Rolls, 2016a). The situation is interestingly analogous in the dorsal visual system, according to the concepts developed here. The successive stages shown in Fig. 1 can be thought of as performing transform invariance computations as described at the start of the Discussion. Layer 1 uses gain modulation of a retinal signal by eye position, and trace learning, to compute representations that are invariant with respect to retinal and eye position, producing head-centered representations. Layer 2 uses gain modulation by head direction and trace learning to compute representations that are invariant with respect to head direction, producing bearing representations to a given stimulus or landmark. Layer 3 uses gain modulation by place and trace learning to compute representations that are invariant with respect to bearing and place (see Fig. 5a), which is important for remembering the location 'out there' where objects or rewards have been seen, and for generalization across different bearings to a location in a scene, and across different places from which the scene is viewed. Both the ventral and dorsal streams can thus be seen to be producing invariant representations, of objects in the ventral stream, and of locations 'out there' in space for the dorsal visual stream. This then helps the hippocampus to associate together invariant representations of objects with invariant representations of locations in a scene 'out there' for primate including human episodic memory. The advantage of this computational design is that once the object-place association has been stored on a single occasion by the hippocampal episodic memory system, the memory can be used later even if the object is seen in a different transform, or the location is seen with a different retinal position, eye position, head direction, allocentric bearing and from a different place. As emphasized earlier, it is important that the information in the dorsal visual system has some selectivity with respect to the part of the scene being viewed, for the bearing to a particular landmark is needed to build that part of a scene, the landmark, into the whole scene, in a way that enables that location in a scene to be represented in a bearing-independent way, which is what is implemented by spatial view cells (Georges-François et al., 1999; Rolls and Wirth, 2018).

Sixth, it has been suggested that gain modulation might be useful in learning for example translation-invariant representations of objects in the ventral visual system (Salinas and Abbott, 1997). However, the results described here show that even if that were the case, then temporal trace learning would be very helpful in allowing neurons to achieve useful and accurate invariance. However, the concept in VisNet is that temporal trace learning is sufficient to learn transform invariant representations of objects in the ventral visual system, with the great advantage that this mechanism can also account for other forms of invariance, including view invariant representations, which cannot be learned by a spatial coordinate transform (as different views of a given object may be completely different), but which VisNet with its temporal short-term memory trace approach learns well (Bart and Hegde, 2012; Perry et al., 2006; Robinson and Rolls, 2015; Rolls, 2012b; Rolls, 2016a; Rolls and Mills, 2018; Rolls and Stringer, 2006; Rolls and Webb, 2014; Wallis and Rolls, 1997; Webb and Rolls, 2014; Zhao et al., 2019).

Seventh, it is argued here that gain modulation of a head-based representation (i.e. the angle of a stimulus or landmark with respect to the head) would be converted by gain modulation by head direction into a coordinate framework of bearing direction to a stimulus or landmark from the animal or agent. This occurs in Layer 2 of the model described here. Neurons of this type may be what has been described in macaque area 7a (Snyder et al., 1998). The population of area 7a neurons was described as responding in a world-based coordinate frame, on the basis that they responded in a particular allocentric direction from the macaque when visually evoked or delayed saccades were made after combined head-and-body rotation in the dark (Snyder et al., 1998). The important point is that when the head was rotated, the area 7a neurons were gain modulated by the head direction. The parsimonious interpretation is that this 'world-based' or allocentric representation is thus in the coordinate frame of bearing direction to a stimulus or landmark. This could facilitate saccade-making in a given bearing direction, that is, independently of the head direction gain modulating factor. Similar results were found by Dean and Platt for the posterior cingulate cortex (which receives inputs from the parietal cortex including area 7a), who showed that for most neurons, tuning

curves aligned more closely when plotted as a function of target position in the room than when plotted as a function of target position with respect to the monkey (Dean and Platt, 2006). A further interesting experiment would be to move the macaque sideways (or the world at which the monkey was looking), to distinguish bearing direction to a landmark in allocentric space from location of a landmark in allocentric space. (That manipulation was in fact used by Feigenbaum and Rolls (1991), and showed that many hippocampal spatial view neurons code for location in allocentric space, with further evidence involving movement of the macaque to different places relative to the location being viewed providing evidence consistent with this (Georges-François et al., 1999; Rolls and O'Mara, 1995). But for now, the evidence for area 7a is that some neurons in it code for bearing direction with respect to the animal, in that the neuronal responses are gain modulated by head direction (Snyder et al., 1998). The requisite signal for this gain modulation is head direction, which is represented by neurons in the primate presubiculum (Robertson et al., 1999).

The important point has been made here that there are several different types of allocentric representation of space, at least in primates. One is an allocentric representation of place, found in the rodent hippocampus (McNaughton et al., 1983; O'Keefe, 1979; O'Keefe and Dostrovsky, 1971) and in the primate hippocampus (Furuya et al., 2014; Hazama and Tamura, 2019; Rolls and O'Mara, 1995; Wirth et al., 2017). A second is an allocentric representation of the location being looked at in the world, as implemented by spatial view neurons in the primate hippocampus (Baraduc et al., 2019; Georges-François et al., 1999; Rolls et al., 1997a; Rolls et al., 1998; Rolls and Wirth, 2018; Wirth and Baraduc, 2018; Wirth et al., 2017). Another is bearing direction, which appears to be implemented by neurons in the primate parietal cortex area 7a (Snyder et al., 1998) and in the primate posterior cingulate cortex (Dean and Platt, 2006). (These bearing direction neurons are different from head direction cells, in that bearing cells to a location are independent of head direction.) Another allocentric frame of reference is head direction, to which neurons in the rat post-subiculum (Cullen and Taube, 2017; Taube et al., 1990) and macaque presubiculum (Robertson et al., 1999) are tuned, thus encoding in compass direction coordinates.

Eighth, although the head-based coordinate frame computed in Layer 1 of the model described here, and the allocentric bearing direction to a location-based frame of reference computed in Layer 2 are supported by current neurophysiological evidence, what is described for Layer 3, transformation of coordinate frames in the dorsal visual system into an allocentric location coordinate frame is, as far as I know, new, and a prediction of the model. The model suggests that the processing in the dorsal visual system for whatever combination of retinal and eye positions, head direction, and place is present, generates representations in an allocentric coordinate framework that can be easily interfaced to hippocampal processing which is in the same allocentric spatial view / spatial location coordinate frame. For example, looking at one location in allocentric space as defined by the current status of the dorsal visual system could provide the allocentric spatial input to the hippocampal memory system via the retrosplenial cortex and/or posterior cingulate cortex (see Fig. 1) for location-object memory retrieval by hippocampal mechanisms including recall of the object from CA3 when an appropriate allocentric spatial location cue is applied. Indeed, something like this is exactly what is proposed to account for the fact that hippocampal spatial view neurons update the allocentric location in space 'out there' to which they respond when eye movements, head direction changes, or even locomotion are made in the dark (Robertson et al., 1998). The idiothetic update could be performed in the dorsal visual system based on the vestibular, proprioceptive, and corollary discharge related signal that reach the dorsal stream visual areas and update spatial representations in it, with examples including the update of representations made for example by eye movements described here, and by vestibular signals (Avila et al., 2019; Chen et al., 2018). Further, in virtual reality, some macaque hippocampal neurons can respond to a view location towards which eye movements are made even before the view has actually appeared on the screen (Wirth et al., 2017). That is potentially adaptive, by speeding up the operation of the system, and can, it is proposed here, be produced by the input from the dorsal visual system (which has information about eye movements etc) but is converted into the correct allocentric view representation by mechanisms of the type described here. Indeed, this empirical evidence is very much in support of the need for a system for coordinate transforms to the allocentric level in the dorsal visual system of the type proposed here. In fact, what the research described here presents is a computational neuronal network based theory and model of some of the coordinate transforms in the dorsal visual system, that enables dorsal visual system representations to be transformed into the same coordinate framework as hippocampal spatial view neurons, to allow simple information transfer between the two systems via areas such as the retrosplenial cortex and posterior cingulate cortex.

Given the ideas just considered on what an allocentric spatial view output of the dorsal visual system might be used for, it is of interest to consider the capacity of the system described here, and how the networks described might scale up with larger numbers of neurons. For a system that operates with discrete representations, the capacity of the system in terms of the number of patterns that can be stored is well defined and can be rigorously calculated and estimated numerically too by simulation (Amit, 1989; Hopfield, 1982; Rolls, 2012a; Treves and Rolls, 1991). However, the capacity for processing analogue information of the type considered here as positions in a continuous spatial representation is harder to estimate, as the precision of the representation then becomes a key factor, but progress has been made by Battaglia and Treves (1998) in their analysis of the capacity of the hippocampus to store spatial charts. Further, in this context, it is unlikely that idiothetic update, one of the functions proposed for the system described here, is very accurate and is implemented in the brain with great precision or over more than a few minutes (Authie et al., 2015; Israel et al., 1997).

Ninth, another issue is how allocentric spatial representations of, for example, spatial location as encoded by spatial view neurons in the primate hippocampus (Rolls, 1999; Rolls and Wirth, 2018; Rolls and Xiang, 2006) are interfaced to egocentric representations of for example arm movements that are made relative to the body frame. For example, some primate hippocampal neurons represent information about the spatial view and the reward or object available at the location being viewed (Baraduc et al., 2019; Rolls and Xiang, 2005; Rolls et al., 2005; Wirth et al., 2009). How would one make a movement to obtain the reward or object that is encoded by memory recall from the primate hippocampus (Rolls and Xiang, 2006)? Major coordinate transforms are required for this. The proposal is that the spatial view neurons respond when the eyes are fixating a particular location in the scene. The primate can then perform arm movements towards the position in space being visually fixated relative to the body, which is in purely egocentric coordinates. That is the type of function in which the parietal cortex specializes (Bremner and Andersen, 2014; Whitlock, 2017). Thus this coordinate transform problem is solved, it is proposed, by the coordinates being effectively passed from the hippocampal allocentric spatial view representation not by a spatial transform computation in the brain, but instead through the world because spatial view cells respond to the location being fixated by the eyes, and the primate can use the position of its eyes, in egocentric coordinates, to perform the arm movement required that is also in egocentric coordinates. This is a major advantage, emphasized here, of the representation of space provided by primate hippocampal spatial view cells (Rolls and Wirth, 2018).

Might similar processes be involved in navigation? If during navigation a primate (including of course human) hippocampal system can use associations between spatial locations and objects or rewards, this would be useful in navigation based on a sequence of landmarks. Remembering the sequence of spatial locations / landmarks (which specify a route) would enable movements towards each of a series of landmarks. Sequence memory may be an important part of hippocampal function (Buzsaki and Tingley, 2018; Eichenbaum, 2017; Foster and Wilson, 2006; Kesner and Rolls, 2015; Lisman and Redish, 2009; Rolls and Mills, 2019). Using the mechanisms described before, the navigation would be by a sequence of remembered allocentric representations: once each remembered allocentric location is viewed and activates spatial view cells, the navigation to the next landmark can be based on the egocentric information about eye and head position that provides the information for the body movements needed to approach the next landmark being viewed. If it is a well learned route, the sequence of spatial views could be retrieved in the correct order using a continuous attractor network that includes spatial views (Stringer et al., 2005). If it is a new route used once or a few times, the sequence could be stored by associating each spatial view or landmark to each step in the sequence of a hippocampal time cell memory (Kraus et al., 2013; MacDonald and Eichenbaum, 2009; Pastalkova et al., 2008; Rolls and Mills, 2019), with analogous mechanisms present in rodents for sequences of places or objects (Buzsaki and Tingley, 2018; Kesner and Rolls, 2015).

If the navigation is in the dark, then it could be performed by idiothetic update of the distance and direction covered, and then remembered body turns after that distance had been travelled. This is a system based primarily on egocentric information, about self-movements, their direction, and body turns (Rolls and Wirth, 2018).

If the navigation is in a maze, the situation may be more complicated. When travelling in one direction in a maze, the spatial view visible from that part of the maze, which is allocentric information, could be associated with body turns that would occur at a particular place, so all these different representations could be used in combinations as the maze is traversed, and linked together in a continuous attractor network (Rolls and Wirth, 2018).

It is of interest to compare the present model with a recent model (Bicanski and Burgess, 2018) centered on the retrosplenial cortex. This previous model is directed primarily towards spatial memory and imagery; does not deal with how eye movements affect spatial vision which is a major part of what is implemented in the primate dorsal visual system in areas such as LIP, VIP, and area 7a together with mechanisms for reaching into space (with eye movements and reaching into space poorly developed and understood in rodents, and no set of highly developed dorsal stream cortical areas in rodents); and relies (Bicanski and Burgess, 2018) on boundary-vector cells found in rodents (with a consistent human fMRI study (Shine et al., 2019)) and not known by neuronal recording evidence to be present in primates. The previous model (Bicanski and Burgess, 2018) also utilizes object-vector cells found in rodents (Hoydal et al., 2019). (Indeed, the description of what Bicanski and Burgess (2018) model is: "Perceived and imagined egocentric sensory experience is represented in the 'parietal window' (PW), which consists of two neural populations - one coding for extended boundaries ('PWb neurons'), and one for discrete objects ('PWo neurons').) Moreover, that model holds that "the transformation between egocentric (parietal) and allocentric (medial temporal lobe, MTL) reference frames is performed by a gain-field circuit in retrosplenial cortex" and uses head direction. In comparison, the present model introduces a new approach to coordinate transforms that includes a memory trace learning rule in competitive networks that is combined with gain modulation for multiple stages of the dorsal visual system; and specifically deals with the spatial transforms and representations known to be implemented in the primate dorsal visual system, and in areas such as the posterior cingulate cortex (not known to be present in rodents). The signals and transforms considered here in comparison include retinal position, eye position, head rotation, and body translation to a different place. The representations that are produced in the research described here are to spatial frameworks with head-centered coordinates, body-centered coordinates for reaching into space, bearing direction to a landmark, and then to allocentric spatial view cells that respond to viewed location in the world independently of the place where the viewer is. Moreover, in the present approach, evidence is provided that the coordinate transform with gain modulation by head direction takes place in parietal areas such as area 7a (Snyder et al., 1998) and is represented in the primate posterior cingulate cortex (Dean and Platt, 2006) (whereas in the model of Bicanski and Burgess (2018) the retrosplenial cortex is emphasised, see their Fig. 1, though the retrosplenial cortex is implicated in memory and navigation (Vann et al., 2009)); holds that this implements bearing direction to a landmark in a spatial scene; and goes on to show that representations in allocentric spatial view coordinates suitable for interfacing to primate allocentric spatial view representations can be produced using gain modulation by place (which was not part of the previous model). A further difference is that in the model of Bicanski and Burgess (2018) the representations produced seem to be of the position of objects in an allocentric space. In contrast, the model described here (which builds on findings in primates) is that in primates, including humans, spatial scenes and locations in them are encoded by hippocampal spatial view cells, and that transform invariant representations of objects (without any spatial properties of what seem to be involved in the putative 'object vector cells' (Bicanski and Burgess, 2018)) can then be associated in the hippocampus with spatial view cell representations of locations in the scene. The result in the present model is of associations between transform-invariant representations of objects, and their place in a scene. This implements what is prototypical of primate and human episodic memory (Kesner and Rolls, 2015; Rolls, 1996; Rolls, 2016a; Rolls and Kesner, 2006; Rolls and Wirth, 2018; Rolls et al., 2005). The transform invariant object and face representations in the primate anterior inferior temporal visual cortex that provide the 'what' input into the hippocampus via the perirhinal cortex have been described elsewhere (DiCarlo et al., 2012; Hong et al., 2016; Rolls, 2012b; Rolls, 2016a; Tsao, 2014).

Overall, the present approach is very different from that of Bicanski and Burgess (2018) as follows. First, the focus here is on the highly developed visual system of primates, with its fovea which provides high acuity over a small field of vision and which is associated with the presence in primates of spatial view cells, which are implicated in memory, spatial function, and navigation (Rolls and Wirth, 2018), and which have not been described in rodents. Second, coordinate transforms are described here that are relevant to the dorsal visual system, highly developed in primates and little developed in rodents, which is important in spatial vision and the control of eye movements which are necessary given the fovea of primates. Third, a new mechanism is introduced here for coordinate transforms throughout the dorsal visual system that combines trace rule slow learning with gain modulation. It is shown here that use of the trace learning rule improves performance greatly, although most of the points made here about coordinate transforms would hold even if trace rule slow learning were not part of the mechanism, and gain modulation was the main

process. Part of the interest of the trace rule slow learning applied to the dorsal visual system is that this draws out a similarity with the ventral visual system, and emphasizes that invariance learning is involved in both the ventral and dorsal visual systems. In the ventral visual system, representations of objects are formed that are partly invariant with respect to transforms such as retinal and eye position, and object size and even view, with trace rule learning implicated (Rolls, 2012b; Rolls, 2016a). In the dorsal visual system, representations of spatial locations "out there", i.e. spatial views, or landmarks, are formed that are partly invariant with respect to transforms such as retinal and eye position, and head direction, and the place where the primate is located, with trace rule learning implicated as described here. Fourth, evidence is provided here that the coordinate transforms take place at different stages of the primate dorsal visual system (and not primarily in the retrosplenial cortex as in the approach of Bicanski and Burgess (2018)). Indeed, as shown in Fig. 1, the retrosplenial cortex is a small region in primates with the much larger posterior cingulate cortex adjoining it (Kobayashi and Amaral, 2003; Rolls, 2019; Vogt, 2009), and the term 'retrosplenial cortex' in rodents may not apply to a closely corresponding area (Vann et al., 2009). The present approach emphasizes the potential importance in coordinate transforms in the primate dorsal visual system of 'bearing to a landmark' (or 'bearing to a spatial view') representations as part of the mechanism for producing an allocentric spatial view representation that is relatively independent of the place from which the landmark is viewed, that is, spatial view cells which are important for remembering where in the environment "out there" objects or rewards are found using the hippocampal memory system (Rolls and Wirth, 2018). However, 'bearing to a landmark' representations may also be important in navigation, as they are one way of establishing the place where one is located, as in the model of de Araujo, Rolls and Stringer (2001). Neurons that might encode 'bearing to a landmark' in primates have been described in parietal cortex area 7a, the posterior cingulate cortex, and hippocampus (Dean and Platt, 2006; Snyder et al., 1998; Wirth et al., 2017), but it will be interesting to explore this much further. (In contrast, Bicanski and Burgess (2018) in their approach rely on cells such as boundary and object vector cells, and primarily on the retrosplenial cortex. A key difference is that Bicanski and Burgess (2018) base their approach on boundary and object vector cells, whereas the new approach described here is based on landmarks / locations in spatial scenes and bearings to these landmarks, which are inherently much more suited to navigation, as well as remembering where one was, or where in a spatial scene or geographically one has seen an object or person.)

In conclusion, the new theory and model presented here provide a powerful general approach to gain modulation computations in the brain, by proposing for the first time that the coordinate transforms and the invariance required could be facilitated by slow temporal trace-related associative learning in a series of dorsal visual stream areas. This is the type of learning implicated in invariance computation in the ventral visual stream (Rolls, 1992; Rolls, 2012b; Rolls, 2016a; Wallis and Rolls, 1997). The research described here shows how the new computational neuronal network-based theory and model of coordinate transforms in the dorsal visual system, could enable dorsal visual system representations to be transformed into the same allocentric location-based in a scene 'out there' coordinate framework as hippocampal spatial view neurons. This may allow simple information transfer between the primate dorsal visual system / parietal cortex and the hippocampal system via areas such as the retrosplenial and posterior cingulate cortex illustrated in Fig. 1. The processes described here show mechanisms by which spatial information from the dorsal visual system of primates could be used to provide a spatial input to the hippocampal memory system in the correct coordinate framework to provide spatial information used in for example object-location memory recall from the hippocampus. This functionality of the dorsal visual system could be important in the idiothetic update of spatial representations. The idiothetically updated spatial representation from the dorsal visual system could then be used as a recall cue for the hippocampal object-location memory system (Kesner and Rolls, 2015; Rolls, 2018), so that what is at a location could be recalled even if that location cannot currently be seen, for example in the dark or when the view is obscured (Robertson et al., 1998), or before the view is even shown towards which the eyes are moving (Rolls and Wirth, 2018; Wirth et al., 2017).

Figure Legends

Fig. 1. Cortical connections of the primate hippocampus. A medial view of the macaque brain is shown below, and a lateral view is above. The entorhinal cortex area 28 is the main entry for cortical connections to and from the hippocampus. The forward projections to the hippocampus are shown with large arrowheads, and the backprojections with small arrowheads. The main ventral stream connections to the hippocampus which convey information about objects, faces, etc are in blue, and the main dorsal stream connections which convey 'where' information about space and movements are in red. The ventral 'what' visual pathways project from the primary visual cortex V1 toV2, then V4, then posterior inferior temporal visual cortex (PIT), then anterior inferior temporal visual cortex (AIT), then perirhinal cortex (areas 35/36), and thus to entorhinal cortex. The dorsal 'where' visual pathways project from V1 to V2, then MT (middle temporal), then LIP (lateral intraparietal), then parietal area 7 (lateral) and medial (including the precuneus), then to posterior cingulate cortex areas 23/32) including the retrosplenial cortex (areas 29/30) and thus to parahippocampal gyrus (areas TF and TH), and then perirhinal and entorhinal cortex. Area 22 is superior temporal auditory association cortex. The hippocampus enables all the high order cortical regions to converge into a single network in the hippocampal CA3 region (Rolls, 2015; Rolls, 2016a). The retrosplenial cortex (29,30) is the small region in primates including humans behind the splenium of the corpus callosum shaded grey: it is not necessarily homologous with what is termed retrosplenial cortex in rodents (Vann et al., 2009), which may also not have a homologous posterior cingulate cortex (Vogt, 2009). Other abbreviations: as–arcuate sulcus; cs–central sulcus; ips–intraparietal sulcus; ios–inferior occipital sulcus; ls–lunate sulcus; sts–superior temporal sulcus. (Modified from Rolls and Wirth 2018.) (HippConnsDV5a.eps)

Fig. 2. The architecture of the VisNetCT model used for the dorsal visual system (see text). Each neuron in a layer (or cortical area in the hierarchy) receives from neurons in a small region of the preceding layer. (VisNetCTArchi2b.eps)

Fig. 3. **Gain modulation by eye position to produce a head-centered representation in Layer 1. a**. Schematic to show that gain modulation by eye position can produce a representation in head-centered coordinates in Layer 1. The eye position (ep) is the angle between straight ahead with respect to the head (indicated by the green line labelled head reference) and the direction of the eye (indicated by the black arrow labelled ep). A direction in head-centered space (labelled hc) is represented by all combinations of retinal position (rp) and eye position that reach a given head-centered direction indicated by the red arrow tip, with one combination of eye position and retinal position shown. **b**. The correlation matrix represents the correlation of the firing between all the neurons in Layer 1 for every one of the five head centred positions (numbered in red) at which a stimulus was presented, after training Layer 1. During training, all combinations of eye position and retinal position that corresponded to one head-centred position were presented together enabling the trace rule to help with the learning, then another head-centred position was selected for training, as described in the Methods. The stimulus order for this diagram was that the first head-centred stimulus position (1 at X=-10, where X refers here to a horizontal plane head-centred position to the left of the head) was the one combination of retinal position and eye position gain modulation signals that corresponded to this head-centred position; the second head-centred stimulus position (2 at X=-5) was the set of two combinations of retinal position and eye position gain modulation signals that corresponded to this head-centred position; the third head-centred stimulus position (3 at X=0) was the set of three combinations of retinal position and eye position gain modulation signals that corresponded to this head-centred position; the fourth head-centred stimulus position (4 at X=+5) was the set of two combinations of retinal position and eye position gain modulation signals that corresponded to this head-centred position; and the fifth head-centred stimulus position (5 at X=+10) was the one combinations of retinal position and eye position gain modulation signals that corresponded to this head-centred position. In the correlation matrix, the large white block in the middle thus indicates that the neurons across the whole of Layer 1 that responded to all three combinations of retinal and eye position for head-centred position 3 encoded only head-centred position 3 for X=0, with no interference from or response to any other retinal and eye position combination that corresponded to other positions in head-centred space. (L2CCmatfinLabgr.eps)

Fig. 4. **Gain modulation by head direction in Layer 2 to produce a representation in bearing coordinates (relative to North) to locations in space at which there are landmarks. a**. Schematic to show that gain modulation by head direction (hd) can produce a representation in bearing coordinates to a stimulus or landmark (L) in Layer 2. The head direction is the angle between North (indicated by the long blue line) and the direction of the head (indicated by the long black arrow). A bearing coordinate to a landmark L is represented by all combinations of head direction, eye position (ep) and retinal position that correspond to a given bearing from the viewer's place to a landmark (L) in allocentric space indicated by the line with a red arrow. Other conventions as in Fig. 3a. **b**. The correlation matrix represents the correlation of the firing between all the neurons in Layer 2 for every for every one of the seven bearing directions each to a different landmark (numbered in red) from the single place of the viewer, after training Layers 1 and 2. There was one combination of retinal position, eye position, and head direction that corresponded to bearing direction to a landmark of X=-15 and +15; three for each of X=-10 and +10; six for X=-5 and +5; and seven combinations for bearing direction 4 at X=0. During training, all combinations of head direction, eye position and retinal position that corresponded to one allocentric bearing to a landmark were presented to enable the trace rule to operate usefully; and each of the other allocentric bearings to a landmark were then trained similarly in turn.   (L3CCmatfinLabgr.eps)

Fig. 5. **Gain modulation by place to produce an allocentric spatial view representation in Layer 3. a**. Schematic to show that gain modulation by place of a bearing representation to a landmark L in Layer 2 can produce a representation of a landmark L in a scene in spatial view coordinates in Layer 3 that is independent of the place where the viewer is located. b1: bearing of a landmark L from place 1; b2: bearing of the same landmark L from place 2; hd1: head direction 1; hd2: head direction 2; ep: eye position; rp: retinal position. A landmark L at a location being viewed in allocentric space, that is, a spatial view, is represented by transforms over all places in Layer 3, building on transforms over head direction learned in Layer 2, and transforms over eye position learned in Layer 1. Other conventions as in Figs 3a and 4a. **b**. The correlation matrix represents the correlation of the firing between all the neurons in Layer 3 for 7 spatial views (numbered in red) presented, after training Layers 1-3. (Results for spatial views of 7 different locations in the scene, each independent of the place where the viewer was located, are shown.) The stimulus order for this diagram was that the first spatial view (i.e. for allocentric location 1 in the scene) (1 at X=-15) was the first set of four stimulus-modulator combinations (shown in the top-left); the second spatial view (i.e. for allocentric location 2 in the scene) (2 at X=-10) was the second set, with ten stimulus-modulator combinations; the third spatial view (i.e. for allocentric location 3 in the scene) (3 at X=-5) was the third set, with 16 stimulus-modulator combinations; the fourth spatial view (4 at X=0) was the fourth set, with 19 stimulus-modulator combinations; the fifth spatial view (5 at X=5) was the fifth set, with 16 stimulus-modulator combinations; and so on until the seventh spatial view (7 at X=15) was the seventh set, with four stimulus-modulator combinations. (Data are not shown for spatial views at X=-20 and +20, as only one combination of retinal and eye position, head direction, and place corresponded to these spatial views.) During training, all combinations of bearing to a landmark, head direction, eye position and retinal position that corresponded to one spatial view of a landmark from one place were presented to enable the trace rule to operate usefully; and each of the other places were then trained similarly in turn. (L4CCmatfinLabgr.eps)

Fig. 6. The responses of a Layer 3 cell selective for spatial view after training. The top left panel shows that the average firing rate of the neuron to every combination of retinal position, eye position, head direction, and place that corresponded to a spatial view at location 5 was 16 spikes/s, and that there was no firing to any other spatial view. The other panels show that the same neuron did not discriminate between place, head direction, or eye position, or retinal position. The bottom left panel shows that the same Layer 3 neuron did not have firing that was specific to a given bearing direction, and the bottom right panel shows that the neuron was not specific to position relative to the head, with the different firing just reflecting how frequently these bearing direction and head centred coordinates occurred when the spatial view at position 5 was being tested. (L4_210b.eps)

Fig. 7. The correlation matrix represents the correlation of the firing between all the neurons in Layer 3 for every spatial view (numbered in red) presented, after training Layers 1-3 without a trace rule, and using only an associative Hebbian learning rule for the competitive networks at each

layer. This should be compared with Fig. 5, in which a short-term memory trace learning rule was used. (L4CCmatfinNoTraceLab.eps)
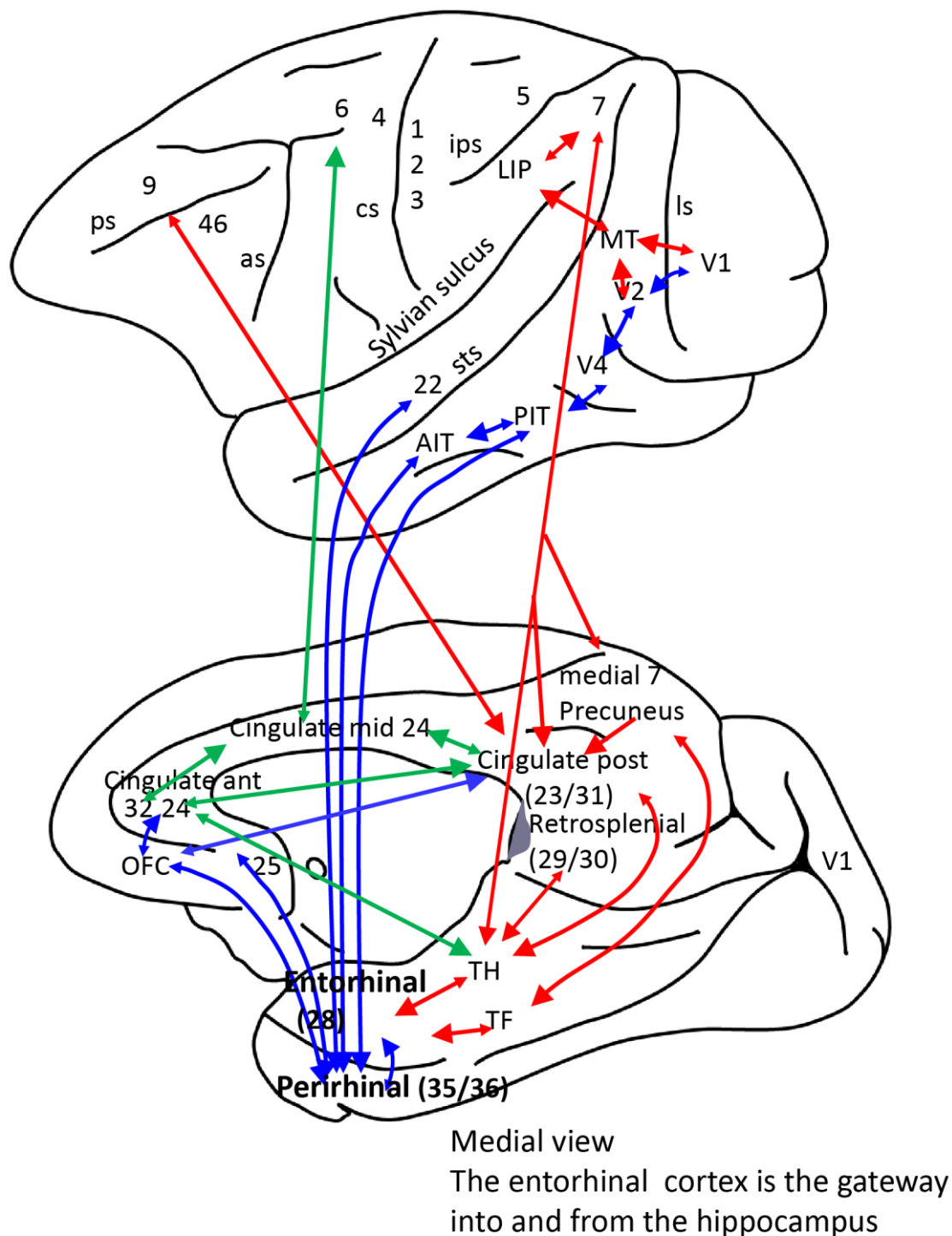

Figure 1
Fig. 1. Cortical connections of the primate hippocampus. A medial view of the macaque brain is shown below, and a lateral view is above. The entorhinal cortex area 28 is the main entry for cortical connections to and from the hippocampus. The forward projections to the hippocampus are shown with large arrowheads, and the backprojections with small arrowheads. The main ventral stream connections to the hippocampus which convey information about objects, faces, etc are in blue, and the main dorsal stream connections which convey 'where' information about space and movements are in red. The ventral 'what' visual pathways project from the primary visual cortex

V1 toV2, then V4, then posterior inferior temporal visual cortex (PIT), then anterior inferior temporal visual cortex (AIT), then perirhinal cortex (areas 35/36), and this to entorhinal cortex. The dorsal 'where' visual pathways project from V1 to V2, then MT (middle temporal), then LIP (lateral intraparietal), then parietal area 7 (lateral) and medial (including the precuneus), then to posterior cingulate cortex areas 23/32) including the retrosplenial cortex (areas 29/30) and thus to parahippocampal gyrus (areas TF and TH), and then perirhinal and entorhinal cortex. Area 22 is superior temporal auditory association cortex. The hippocampus enables all the high order cortical regions to converge into a single network in the hippocampal CA3 region (Rolls, 2015; Rolls, 2016a). The retrosplenial cortex (29,30) is the small region in primates including humans behind the splenium of the corpus callosum shaded grey: it is not necessarily homologous with what is termed retrosplenial cortex in rodents (Vann et al., 2009), which may also not have a homologous posterior cingulate cortex (Vogt, 2009). Other abbreviations: as–arcuate sulcus; cs–central sulcus; ips–intraparietal sulcus; ios–inferior occipital sulcus; ls–lunate sulcus; sts–superior temporal sulcus. (Modified from Rolls and Wirth 2018.) (HippConnsDV5a.eps)
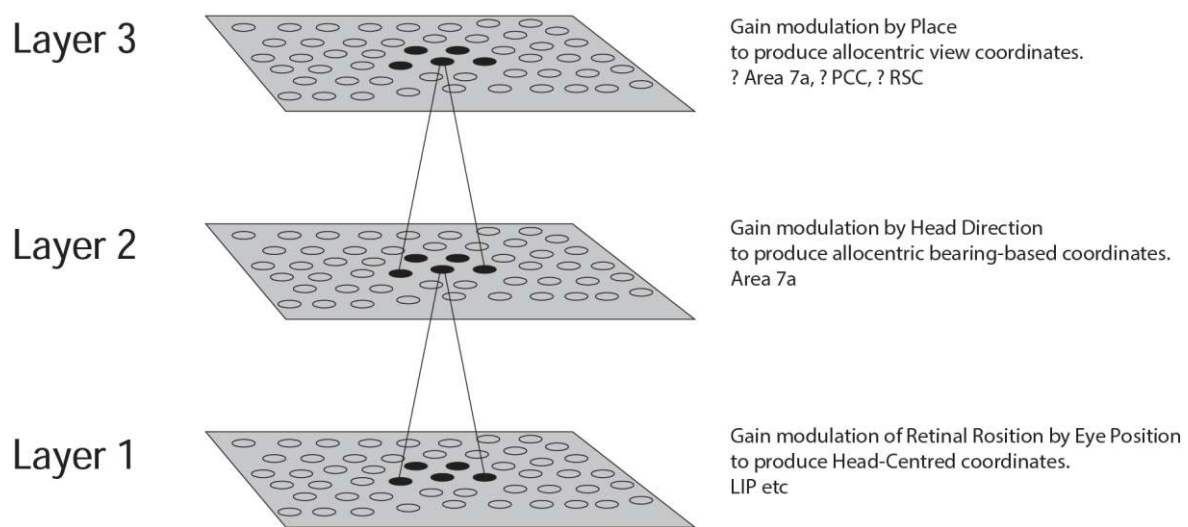
Fig. 2. The architecture of the VisNetCT model used for the dorsal visual system (see text). Each neuron in a layer (or cortical area in the hierarchy) receives from neurons in a small region of the preceding layer. PCC - posterior cingulate cortex; RSC - retrosplenial cortex. (VisNetCTArchi2b.eps)
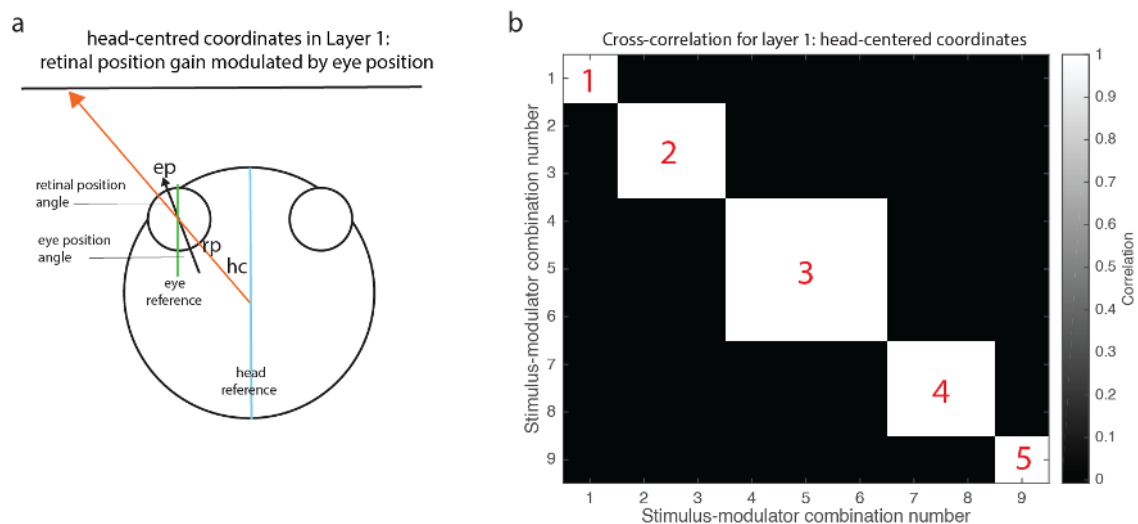
Fig. 3. **Gain modulation by eye position to produce a head-centered representation in Layer 1. a**. Schematic to show that gain modulation by eye position can produce a representation in head-centered coordinates in Layer 1. The eye position (ep) is the angle between straight ahead with respect to the head (indicated by the green line labelled head reference) and the direction of the eye (indicated by the black arrow labelled ep). A direction in head-centered space (labelled hc) is represented by all combinations of retinal position (rp) and eye position that reach a given head-centered direction indicated by the red arrow tip, with one combination of eye position and retinal position shown. **b**. The correlation matrix represents the correlation of the firing between all the neurons in Layer 1 for every one of the five head centred positions (numbered in red) at which a stimulus was presented, after training Layer 1. During training, all combinations of eye position and retinal position that corresponded to one head-centred position were presented together enabling the trace rule to help with the learning, then another head-centred position was selected for graining, as described in the Methods. The stimulus order for this diagram was that the first head-centred stimulus position (1 at X=-10, where X refers here to a head-centred position to the left of the head) was the one combination of retinal position and eye position gain modulation signals that corresponded to this head-centred position; the second head-centred stimulus position (2 at X=-5) was the set of two combinations of retinal position and eye position gain modulation signals that corresponded to this head-centred position; the third head-centred stimulus position (3 at X=0) was the set of three combinations of retinal position and eye position gain modulation signals that corresponded to this head-centred position; the fourth head-centred stimulus position (4 at X=+5) was the set of two combinations of retinal position and eye position gain modulation signals that corresponded to this head-centred position; and the fifth head-centred stimulus position (5 at X=-10) was the one combinations of retinal position and eye position gain modulation signals that corresponded to this head-centred position. In the correlation matrix, the large white block in the middle thus indicates that the neurons across the whole of Layer 1 that responded to all three combinations of retinal and eye position for head-centred position 3 encoded only head-centred position 3 for X=0, with no interference from or response to any other retinal and eye position combination that corresponded to other positions in head-centred space. (L2CCmatfinLabgr.eps)
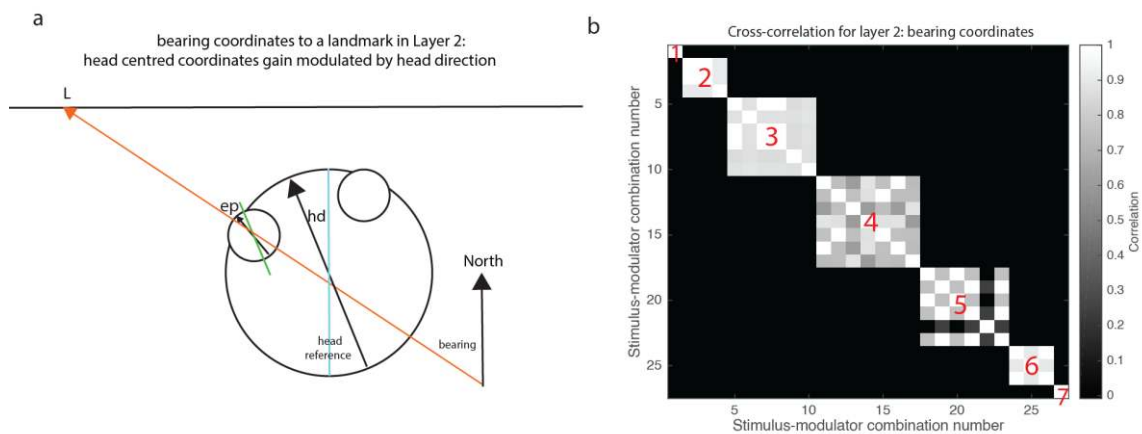
Fig. 4. **Gain modulation by head direction to produce a representation in bearing coordinates (relative to North) to a location in space at which there is a landmark in Layer 2.** **a**. Schematic to show that gain modulation by head direction (hd) can produce a representation in bearing coordinates to a stimulus or landmark (L) in Layer 2. The head direction is the angle between North (indicated by the long blue line) and the direction of the head (indicated by the long black arrow). A bearing coordinate to a landmark L is represented by all combinations of head direction, eye position (ep) and retinal position that correspond to a given bearing from the individual to a landmark in allocentric space indicated by the line with a red arrow. Other conventions as in Fig. 3a. **b**. The correlation matrix represents the correlation of the firing between all the neurons in Layer 2 for every for every one of the seven bearing directions (numbered in red) relative to the agent at which a stimulus was presented, after training Layers 1 and 2. There was one combination of retinal position, eye position, and head direction that corresponded to bearing direction of X=-15 and +15; three for each of X=-10 and +10; six for X=-5 and +5; and seven combinations for bearing direction 4 at X=0. During training, all combinations of head direction, eye position and retinal position that corresponded to one allocentric bearing to a landmark were presented to enable the trace rule to operate usefully; and each of the other allocentric bearings to a landmark were then trained similarly in turn. (L3CCmatfinLabgr.eps)
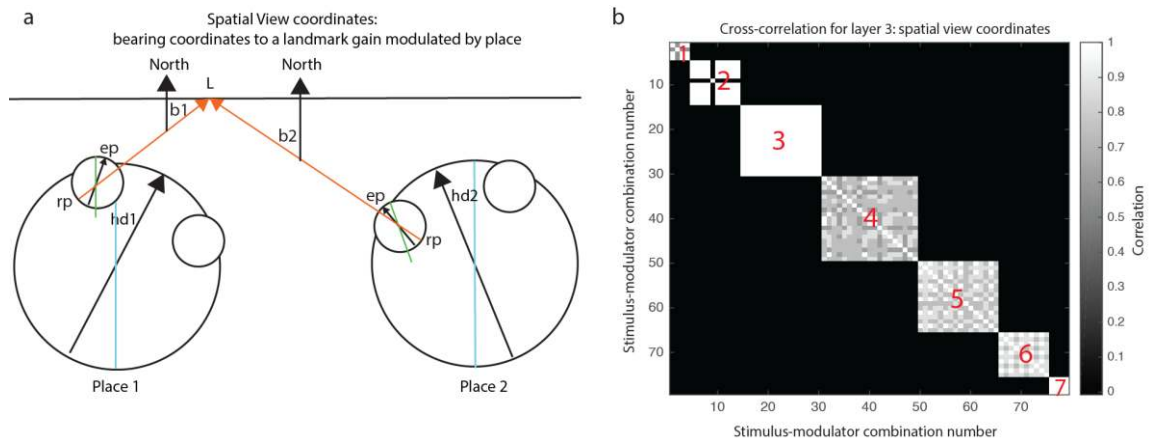
Fig. 5. **Gain modulation by place to produce an allocentric spatial view representation in Layer 3. a**. Schematic to show that gain modulation by place of a bearing representation from Layer 2 can produce a representation of a landmark L in a scene in spatial view coordinates in Layer 3. b1: bearing of the landmark from place 1; b2: bearing of the landmark from place 2; hd1: head direction 1; hd2: head direction 2; ep: eye position; rp: retinal position. A landmark L at a location being viewed in allocentric space, that is, a spatial view, is represented by transforms over all places in Layer 3, building on transforms over head direction learned in Layer 2, and transforms over eye position learned in Layer 1. Other conventions as in Figs 3a and 4a. **b**. The correlation matrix represents the correlation of the firing between all the neurons in Layer 3 for 7 spatial views (numbered in red) presented, after training Layers 1-3. The stimulus order for this diagram was that the first spatial view (1 at X=-15) was the first set of four stimulus-modulator combinations (shown in the top-left); the second spatial view (2 at X=-10) was the second set, with ten stimulus-modulator combinations; the third spatial view (3 at X=-5) was the third set, with 16 stimulus-modulator combinations; the fourth spatial view (4 at X=0) was the fourth set, with 19 stimulus-modulator combinations; the fifth spatial view (5 at X=5) was the fifth set, with 16 stimulus-modulator combinations; and so on until the seventh spatial view (7 at X=15) was the seventh set, with four stimulus-modulator combinations. The first spatial view (for X= -15) was represented by four combinations of the retinal position and modulators of eye position, head direction, and place), and formed one block of correlated representations in the top left of Fig. 5a. (Data are not shown for spatial views at X=-20 and +20, as only one combination of retinal and eye position, head direction, and place corresponded to these spatial views.) During training, all combinations of bearing to a landmark, head direction, eye position and retinal position that corresponded to one spatial view of a landmark from one place were presented to enable the trace rule to operate usefully; and each of the other places were then trained similarly in turn. (L4CCmatfinLabgr.eps)
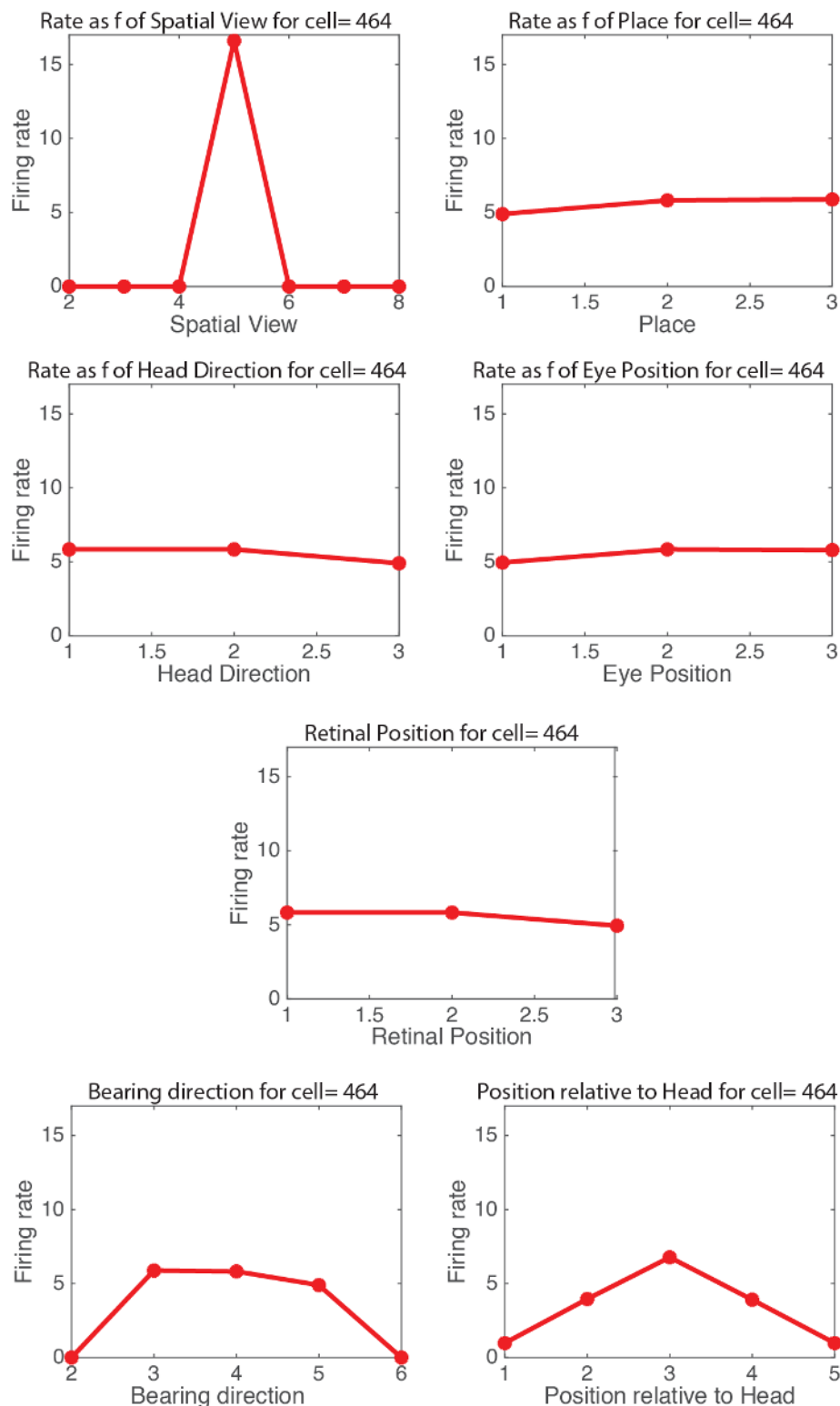
Fig. 6

Fig. 6. The responses of a Layer 3 cell selective for spatial view after training. The top left panel shows that the average firing rate of the neuron to every combination of retinal position, eye position, head direction, and place that corresponded to a spatial view at location 5 was 16 spikes/s, and that there was no firing to any other spatial view. The other panels show that the same neuron did not discriminate between place, head direction, or eye position, or retinal position. The bottom left panel shows that the same Layer 3 neuron did not have firing that was specific to a given bearing direction, and the bottom right panel shows that the neuron was not specific to position relative to the head, with the different firing just reflecting how frequently these bearing direction and head centred coordinates occurred when the spatial view at position 5 was being tested. (L4_210b.eps)
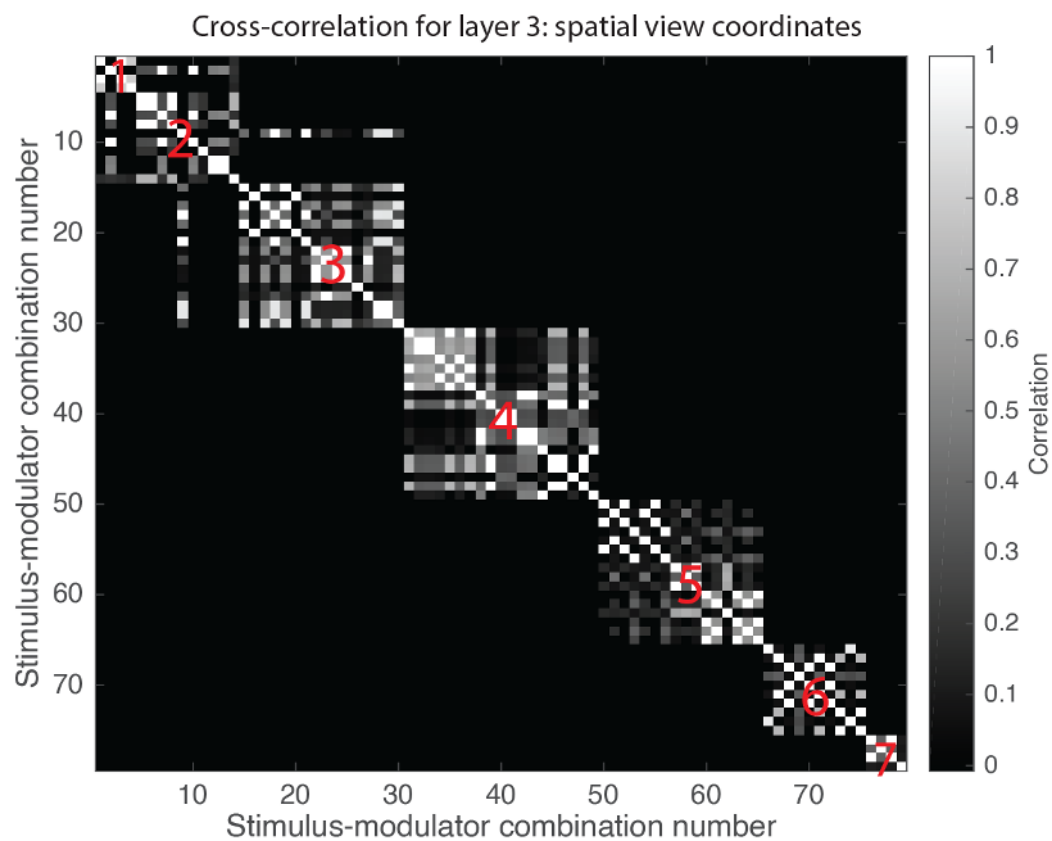
Fig. 7. The correlation matrix represents the correlation of the firing between all the neurons in Layer 3 for every spatial view (numbered in red) presented, after training Layers 1-3 without a trace rule, and using only an associative Hebbian learning rule for the competitive networks at each layer. This should be compared with Fig. 5, in which a short-term memory trace learning rue was used. (L4CCmatfinNoTraceLab.eps)

Table 1. The default parameters used in the VisNetCT model, unless otherwise stated.

Neurons in each layer	32x32
Number of synapses onto each neuron	100
Radius of Gaussian input connectivity of each neuron	$\sigma = 2$
η the trace update proportion	0.8
Learning rate α	0.05
Sparseness a in each layer	0.008
Number of retinal positions (-5 0 5)	3
Number of eye positions (-5 0 5)	3
Number of head directions (-5 0 5)	3
Number of places (-5 0 5)	3
Size of retinal stimulus: 1 pixel convolved with a kernel with $\sigma = 1$	
Number of training epochs for each layer	12

References

Aguirre GK, D'Esposito M. 1999. Topographical disorientation: a synthesis and taxonomy. Brain 122 (Pt 9):1613-28.

Alexander AS, Nitz DA. 2015. Retrosplenial cortex maps the conjunction of internal and external spaces. Nature Neuroscience 18(8):1143-1151.

Amit DJ. 1989. Modeling Brain Function. Cambridge: Cambridge University Press.

Andersen RA. 1989. Visual and eye movement functions of the posterior parietal cortex. Annu Rev Neurosci 12:377-403.

Andersen RA, Cui H. 2009. Intention, action planning, and decision making in parietal-frontal circuits. Neuron 63(5):568-83.

Andersen RA, Essick GK, Siegel RM. 1985. Encoding of spatial location by posterior parietal neurons. Science 230(4724):456-8.

Andersen RA, Mountcastle VB. 1983. The influence of the angle of gaze upon the excitability of the light-sensitive neurons of the posterior parietal cortex. J Neurosci 3(3):532-48.

Authie CN, Hilt PM, N'Guyen S, Berthoz A, Bennequin D. 2015. Differences in gaze anticipation for locomotion with and without vision. Front Hum Neurosci 9:312.

Avila E, Lakshminarasimhan KJ, DeAngelis GC, Angelaki DE. 2019. Visual and vestibular selectivity for self-motion in macaque Posterior Parietal Area 7a. Cereb Cortex 29:3932-3947.

Baraduc P, Duhamel JR, Wirth S. 2019. Schema cells in the macaque hippocampus. Science 363(6427):635-639.

Bart E, Hegde J. 2012. Invariant recognition of visual objects: some emerging computational principles. Front Comput Neurosci 6:60.

Barton JJ. 2011. Disorders of higher visual processing. Handb Clin Neurol 102:223-61.

Battaglia FP, Treves A. 1998. Attractor neural networks storing multiple space representations: a model for hippocampal place fields. Physical Review E 58:7738-7753.

Bicanski A, Burgess N. 2018. A neural-level model of spatial memory and imagery. Elife 7:e33752.

Bohbot VD, Corkin S. 2007. Posterior parahippocampal place learning in H.M. Hippocampus 17(9):863-72.

Bremmer F, Klam F, Duhamel JR, Ben Hamed S, Graf W. 2002. Visual-vestibular interactive responses in the macaque ventral intraparietal area (VIP). Eur J Neurosci 16(8):1569-86.

Bremner LR, Andersen RA. 2014. Temporal analysis of reference frames in parietal cortex area 5d during reach planning. J Neurosci 34(15):5273-84.

Burgess N, Hartley T. 2001. Orietational and geometric determinants of place and head-direction. Advances in Neural Information Processing Systems 14(2):165-172.

Buzsaki G, Tingley D. 2018. Space and Time: The Hippocampus as a Sequence Generator. Trends Cogn Sci 22(10):853-869.

Byrne P, Becker S, Burgess N. 2007. Remembering the past and imagining the future: a neural model of spatial memory and imagery. Psychol Rev 114(2):340-75.

Chen X, DeAngelis GC, Angelaki DE. 2018. Flexible egocentric and allocentric representations of heading signals in parietal cortex. Proc Natl Acad Sci U S A 115(14):E3305-E3312.

Clark IA, Maguire EA. 2016. Remembering preservation in hippocampal amnesia. Annu Rev Psychol 67:51-82.

Cullen KE, Taube JS. 2017. Our sense of direction: progress, controversies and challenges. Nat Neurosci 20(11):1465-1473.

de Araujo IET, Rolls ET, Stringer SM. 2001. A view model which accounts for the spatial fields of hippocampal primate spatial view cells and rat place cells. Hippocampus 11:699-706.

Dean HL, Platt ML. 2006. Allocentric spatial referencing of neuronal activity in macaque posterior cingulate cortex. J Neurosci 26(4):1117-27.

Deshmukh SS, Knierim JJ. 2013. Influence of local objects on hippocampal representations: Landmark vectors and memory. Hippocampus 23(4):253-67.

DiCarlo JJ, Zoccolan D, Rust NC. 2012. How does the brain solve visual object recognition? Neuron 73(3):415-34.

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

Eguchi A, Humphreys GW, Stringer SM. 2016. The visually guided development of facial representations in the primate ventral visual pathway: A computer modeling study. Psychol Rev 123(6):696-739.

Eichenbaum H. 2017. On the Integration of Space, Time, and Memory. Neuron 95(5):1007-1018.

Ekstrom AD, Arnold AE, Iaria G. 2014. A critical review of the allocentric spatial representation and its neural underpinnings: toward a network-based perspective. Front Hum Neurosci 8:803.

Ekstrom AD, Huffman DJ, Starrett M. 2017. Interacting networks of brain regions underlie human spatial navigation: a review and novel synthesis of the literature. J Neurophysiol 118(6):3328-3344.

Epstein RA. 2008. Parahippocampal and retrosplenial contributions to human spatial navigation. Trends Cogn Sci 12(10):388-96.

Feigenbaum JD, Rolls ET. 1991. Allocentric and egocentric spatial information processing in the hippocampal formation of the behaving primate. Psychobiology 19:21-40.

Földiák P. 1991. Learning invariance from transformation sequences. Neural Computation 3(2):194-200.

Foster DJ, Wilson MA. 2006. Reverse replay of behavioural sequences in hippocampal place cells during the awake state. Nature 440(7084):680-3.

Franco L, Rolls ET, Aggelopoulos NC, Jerez JM. 2007. Neuronal selectivity, population sparseness, and ergodicity in the inferior temporal visual cortex. Biological Cybernetics 96(6):547-560.

Franzius M, Sprekeler H, Wiskott L. 2007. Slowness and sparseness lead to place, head-direction, and spatial-view cells. PLoS Comput Biol 3(8):e166.

Furuya Y, Matsumoto J, Hori E, Boas CV, Tran AH, Shimada Y, Ono T, Nishijo H. 2014. Place-

related neuronal activity in the monkey parahippocampal gyrus and hippocampal formation during virtual navigation. Hippocampus 24(1):113-30.

Galletti C, Fattori P. 2018. The dorsal visual stream revisited: Stable circuits or dynamic pathways? Cortex 98:203-217.

Georges-François P, Rolls ET, Robertson RG. 1999. Spatial view cells in the primate hippocampus: allocentric view not head direction or eye position or place. Cerebral Cortex 9:197-212.

Gold AE, Kesner RP. 2005. The role of the CA3 subregion of the dorsal hippocampus in spatial pattern completion in the rat. Hippocampus 15:808-814.

Graf AB, Andersen RA. 2014. Inferring eye position from populations of lateral intraparietal neurons. Elife 3:e02813.

Habib M, Sirigu A. 1987. Pure topographical disorientation: a definition and anatomical basis. Cortex 23:73-85.

Hazama Y, Tamura R. 2019. Effects of self-locomotion on the activity of place cells in the hippocampus of a freely behaving monkey. Neurosci Lett 701:32-37.

Hong H, Yamins DL, Majaj NJ, DiCarlo JJ. 2016. Explicit information for category-orthogonal object properties increases along the ventral stream. Nat Neurosci 19(4):613-22.

Hopfield JJ. 1982. Neural networks and physical systems with emergent collective computational abilities. Proceedings of the National Academy of Science USA 79:2554-2558.

Hoydal OA, Skytoen ER, Andersson SO, Moser MB, Moser EI. 2019. Object-vector coding in the medial entorhinal cortex. Nature 568(7752):400-404.

Israel I, Grasso R, Georges-Francois P, Tsuzuku T, Berthoz A. 1997. Spatial memory and path integration studied by self-driven passive linear displacement. I. Basic properties. J Neurophysiol 77(6):3180-92.

Kesner RP, Hunsaker MR, Warthen MW. 2008. The CA3 subregion of the hippocampus is critical for episodic memory processing by means of relational encoding in rats. Behavioral Neuroscience 122(6):1217-1225.

Kesner RP, Rolls ET. 2015. A computational theory of hippocampal function, and tests of the theory: new developments. Neuroscience and Biobehavioral Reviews 48:92-147.

Killian NJ, Jutras MJ, Buffalo EA. 2012. A map of visual space in the primate entorhinal cortex. Nature 491(7426):761-4.

Kim JG, Aminoff EM, Kastner S, Behrmann M. 2015. A Neural Basis for Developmental Topographic Disorientation. J Neurosci 35(37):12954-69.

Kobayashi Y, Amaral DG. 2003. Macaque monkey retrosplenial cortex: II. Cortical afferents. J Comp Neurol 466:48-79.

Kobayashi Y, Amaral DG. 2007. Macaque monkey retrosplenial cortex: III. Cortical efferents. J Comp Neurol 502(5):810-33.

Kolb B, Whishaw IQ. 2015. Fundamentals of Human Neuropsychology. New York: Worth.

Kornblith S, Cheng X, Ohayon S, Tsao DY. 2013. A network for scene processing in the macaque temporal lobe. Neuron 79(4):766-81.

Kraus BJ, Robinson RJ, White JA, Eichenbaum H, Hasselmo ME. 2013. Hippocampal "time

cells": time versus path integration. Neuron 78(6):1090-1101.

LeCun Y, Bengio Y, Hinton G. 2015. Deep learning. Nature 521(7553):436-44.

Lisman J, Redish AD. 2009. Prediction, sequences and the hippocampus. Philos Trans R Soc Lond B Biol Sci 364(1521):1193-201.

MacDonald CJ, Eichenbaum H. 2009. Hippocampal neurons disambiguate overlapping sequences of non-spatial events. Society for Neuroscience Abstracts 101.21.

Maguire EA. 2001. The retrosplenial contribution to human navigation: a review of lesion and neuroimaging findings. Scand J Psychol 42(3):225-38.

Maguire EA, Intraub H, Mullally SL. 2016. Scenes, spaces, and memory traces: what does the hippocampus do? Neuroscientist 22(5):432-9.

Marr D. 1971. Simple memory: a theory for archicortex. Phil. Trans. Roy. Soc. Lond. B 262:23-81.

McNaughton BL, Barnes CA, O'Keefe J. 1983. The contributions of position, direction, and velocity to single unit activity in the hippocampus of freely-moving rats. Exp Brain Res 52(1):41-9.

McNaughton BL, Morris RGM. 1987. Hippocampal synaptic enhancement and information storage within a distributed memory system. Trends in Neurosciences 10(10):408-415.

Meister MLR, Buffalo EA. 2018. Neurons in Primate Entorhinal Cortex Represent Gaze Position in Multiple Spatial Reference Frames. J Neurosci 38(10):2430-2441.

Miller JF, Neufang M, Solway A, Brandt A, Trippel M, Mader I, Hefft S, Merkow M, Polyn SM, Jacobs J and others. 2013. Neural activity in human hippocampal formation reveals the spatial context of retrieved memories. Science 342(6162):1111-4.

Moser EI, Moser MB. 2003. One-shot memory in hippocampal CA3 networks. Neuron 38(2):147-8.

Nakazawa K, Quirk MC, Chitwood RA, Watanabe M, Yeckel MF, Sun LD, Kato A, Carr CA, Johnston D, Wilson MA and others. 2002. Requirement for hippocampal CA3 NMDA receptors in associative memory recall. Science 297:211-218.

Nasr S, Liu N, Devaney KJ, Yue X, Rajimehr R, Ungerleider LG, Tootell RB. 2011. Scene-selective cortical regions in human and nonhuman primates. J Neurosci 31(39):13771-85.

Navarro DM, Mender BMW, Smithson HE, Stringer SM. 2018. Self-organising coordinate transformation with peaked and monotonic gain modulation in the primate dorsal visual pathway. PLoS One 13(11):e0207961.

O'Keefe J. 1979. A review of the hippocampal place cells. Progress in Neurobiology 13:419-439.

O'Keefe J, Dostrovsky J. 1971. The hippocampus as a spatial map: preliminary evidence from unit activity in the freely moving rat. Brain Research 34:171-175.

Oja E. 1982. A simplified neuron model as a principal component analyser. Journal of Mathematical Biology 15:267-273.

Pastalkova E, Itskov V, Amarasingham A, Buzsaki G. 2008. Internally generated cell assembly sequences in the rat hippocampus. Science 321(5894):1322-7.

Perry G, Rolls ET, Stringer SM. 2006. Spatial vs temporal continuity in view invariant visual object recognition learning. Vision Research 46:3994-4006.

Pouget A, Sejnowski TJ. 1997. Spatial transformations in the parietal cortex using basis functions. J Cogn Neurosci 9(2):222-37.

Roberts JA, Wallis G, Breakspear M. 2013. Fixational eye movements during viewing of dynamic natural scenes. Front Psychol 4:797.

Robertson RG, Rolls ET, Georges-François P. 1998. Spatial view cells in the primate hippocampus: Effects of removal of view details. Journal of Neurophysiology 79:1145-1156.

Robertson RG, Rolls ET, Georges-François P, Panzeri S. 1999. Head direction cells in the primate pre-subiculum. Hippocampus 9:206-219.

Robinson L, Rolls ET. 2015. Invariant visual object recognition: biologically plausible approaches. Biological Cybernetics 109(4-5):505-35.

Rolls ET. 1987. Information representation, processing and storage in the brain: analysis at the single neuron level. In: Changeux J-P, Konishi M, editors. The Neural and Molecular Bases of Learning. Chichester: Wiley. p 503-540.

Rolls ET. 1989a. Functions of neuronal networks in the hippocampus and neocortex in memory. In: Byrne JH, Berry WO, editors. Neural Models of Plasticity: Experimental and Theoretical Approaches. San Diego: Academic Press. p 240-265.

Rolls ET. 1989b. The representation and storage of information in neuronal networks in the primate cerebral cortex and hippocampus. In: Durbin R, Miall C, Mitchison G, editors. The Computing Neuron. Wokingham, England: Addison-Wesley. p 125-159.

Rolls ET. 1992. Neurophysiological mechanisms underlying face processing within and beyond the temporal cortical visual areas. Philosophical Transactions of the Royal Society of London B 335:11-21.

Rolls ET. 1996. A theory of hippocampal function in memory. Hippocampus 6:601-620.

Rolls ET. 1999. Spatial view cells and the representation of place in the primate hippocampus. Hippocampus 9:467-480.

Rolls ET. 2012a. Advantages of dilution in the connectivity of attractor networks in the brain. Biologically Inspired Cognitive Architectures 1:44-54.

Rolls ET. 2012b. Invariant visual object and face recognition: neural and computational bases, and a model, VisNet. Frontiers in Computational Neuroscience 6, 35:1-70.

Rolls ET. 2015. Limbic systems for emotion and for memory, but no single limbic system. Cortex 62:119-157.

Rolls ET. 2016a. Cerebral Cortex: Principles of Operation. Oxford: Oxford University Press.

Rolls ET. 2016b. Pattern separation, completion, and categorisation in the hippocampus and neocortex. Neurobiology of Learning and Memory 129:4-28.

Rolls ET. 2018. The storage and recall of memories in the hippocampo-cortical system. Cell and Tissue Research 373:577-604.

Rolls ET. 2019. The cingulate cortex and limbic systems for emotion, action, and memory. Brain Struct Funct.

Rolls ET, Kesner RP. 2006. A computational theory of hippocampal function, and empirical tests of the theory. Progress in Neurobiology 79:1-48.

Rolls ET, Mills P. 2019. The generation of time in the hippocampal memory system. Cell Rep

28(7):1649-1658 e6.

Rolls ET, Mills WPC. 2018. Non-accidental properties, metric invariance, and encoding by neurons in a model of ventral stream visual object recognition, VisNet. Neurobiology of Learning and Memory 152:20-31.

Rolls ET, Milward T. 2000. A model of invariant object recognition in the visual system: learning rules, activation functions, lateral inhibition, and information-based performance measures. Neural Computation 12:2547-2572.

Rolls ET, Miyashita Y, Cahusac PMB, Kesner RP, Niki H, Feigenbaum J, Bach L. 1989. Hippocampal neurons in the monkey with activity related to the place in which a stimulus is shown. Journal of Neuroscience 9:1835-1845.

Rolls ET, O'Mara SM. 1995. View-responsive neurons in the primate hippocampal complex. Hippocampus 5(5):409-24.

Rolls ET, Robertson RG, Georges-François P. 1997a. Spatial view cells in the primate hippocampus. European Journal of Neuroscience 9:1789-1794.

Rolls ET, Stringer SM. 2001. Invariant object recognition in the visual system with error correction and temporal difference learning. Network: Computation in Neural Systems 12:111-129.

Rolls ET, Stringer SM. 2006. Invariant global motion recognition in the dorsal visual system: a unifying theory. Neural Computation 19:139-169.

Rolls ET, Stringer SM, Elliot T. 2006. Entorhinal cortex grid cells can map to hippocampal place cells by competitive learning. Network: Computation in Neural Systems 17:447-465.

Rolls ET, Tovee MJ. 1995. Sparseness of the neuronal representation of stimuli in the primate temporal visual cortex. Journal of Neurophysiology 73(2):713-726.

Rolls ET, Treves A. 1990. The relative advantages of sparse versus distributed encoding for associative neuronal networks in the brain. Network 1:407-421.

Rolls ET, Treves A. 2011. The neuronal encoding of information in the brain. Progress in Neurobiology 95(3):448-490.

Rolls ET, Treves A, Robertson RG, Georges-François P, Panzeri S. 1998. Information about spatial view in an ensemble of primate hippocampal cells. Journal of Neurophysiology 79:1797-1813.

Rolls ET, Treves A, Tovee MJ. 1997b. The representational capacity of the distributed encoding of information provided by populations of neurons in the primate temporal visual cortex. Experimental Brain Research 114(1):177-185.

Rolls ET, Treves A, Tovee MJ, Panzeri S. 1997c. Information in the neuronal representation of individual stimuli in the primate temporal visual cortex. Journal of Computational Neuroscience 4(4):309-333.

Rolls ET, Webb TJ. 2014. Finding and recognising objects in natural scenes: complementary computations in the dorsal and ventral visual systems. Frontiers in Computational Neuroscience 8:85.

Rolls ET, Wirth S. 2018. Spatial representations in the primate hippocampus, and their functions in memory and navigation. Prog Neurobiol 171:90-113.

Rolls ET, Xiang J-Z. 2005. Reward-spatial view representations and learning in the hippocampus. Journal of Neuroscience 25:6167– 6174.

Rolls ET, Xiang J-Z. 2006. Spatial view cells in the primate hippocampus, and memory recall. Reviews in the Neurosciences 17(1-2):175-200.

Rolls ET, Xiang J-Z, Franco L. 2005. Object, space and object-space representations in the primate hippocampus. Journal of Neurophysiology 94(1):833-844.

Rumelhart DE, Zipser D. 1985. Feature discovery by competitive learning. Cognitive Science 9:75--112.

Salinas E, Abbott LF. 1995. Transfer of coded information from sensory to motor networks. J Neurosci 15(10):6461-74.

Salinas E, Abbott LF. 1996. A model of multiplicative neural responses in parietal cortex. Proc Natl Acad Sci U S A 93(21):11956-61.

Salinas E, Abbott LF. 1997. Invariant visual responses from attentional gain fields. J Neurophysiol 77(6):3267-72.

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

Salinas E, Sejnowski TJ. 2001. Gain modulation in the central nervous system: where behavior, neurophysiology, and computation meet. Neuroscientist 7(5):430-40.

Sarel A, Finkelstein A, Las L, Ulanovsky N. 2017. Vectorial representation of spatial goals in the hippocampus of bats. Science 355(6321):176-180.

Shannon CE. 1948. A mathematical theory of communication. AT&T Bell Laboratories Technical Journal 27:379-423.

Shine JP, Valdes-Herrera JP, Tempelmann C, Wolbers T. 2019. Evidence for allocentric boundary and goal direction information in the human entorhinal cortex and subiculum. Nat Commun 10(1):4004.

Snyder LH, Grieve KL, Brotchie P, Andersen RA. 1998. Separate body- and world-referenced representations of visual space in parietal cortex. Nature 394(6696):887-91.

Spiers HJ, Maguire EA. 2006. Thoughts, behaviour, and brain dynamics during navigation in the real world. Neuroimage 31(4):1826-40.

Stringer SM, Rolls ET, Trappenberg TP. 2005. Self-organizing continuous attractor network models of hippocampal spatial view cells. Neurobiol Learn Mem 83(1):79-92.

Takahashi N, Kawamura M, Shiota J, Kasahata N, Hirayama K. 1997. Pure topographic disorientation due to right retrosplenial lesion. Neurology 49(2):464-9.

Taube JS, Muller RU, Ranck JB, Jr. 1990. Head-direction cells recorded from the postsubiculum in freely moving rats. I. Description and quantitative analysis. J Neurosci 10(2):420-35.

Teng E, Squire LR. 1999. Memory for places learned long ago is intact after hippocampal damage. Nature 400(6745):675-7.

Treves A. 1991. Dilution and sparse coding in threshold-linear nets. Journal of Physics A 24:327-335.

Treves A, Rolls ET. 1991. What determines the capacity of autoassociative memories in the brain? Network 2:371-397.

Treves A, Rolls ET. 1992. Computational constraints suggest the need for two distinct input systems to the hippocampal CA3 network. Hippocampus 2:189-199.

Treves A, Rolls ET. 1994. A computational analysis of the role of the hippocampus in memory. Hippocampus 4:374-391.

Tsao D. 2014. The Macaque Face Patch System: A Window into Object Representation. Cold Spring Harb Symp Quant Biol 79:109-14.

Ungerleider LG, Haxby JV. 1994. 'What' and 'where' in the human brain. Curr Opin Neurobiol 4(2):157-65.

Vann SD, Aggleton JP, Maguire EA. 2009. What does the retrosplenial cortex do? Nat Rev Neurosci 10(11):792-802.

Vedder LC, Miller AMP, Harrison MB, Smith DM. 2017. Retrosplenial Cortical Neurons Encode Navigational Cues, Trajectories and Reward Locations During Goal Directed Navigation. Cereb Cortex 27(7):3713-3723.

Vogt BA, editor. 2009. Cingulate Neurobiology and Disease. Oxford: Oxford University Press.

von der Malsburg C. 1973. Self-organization of orientation sensitive cells in the striate cortex. Kybernetik 14(2):85-100.

Wallis G, Rolls ET. 1997. Invariant face and object recognition in the visual system. Progress in Neurobiology 51:167-194.

Wallis G, Rolls ET, Földiák P. 1993. Learning invariant responses to the natural transformations of objects. International Joint Conference on Neural Networks 2:1087-1090.

Wang C, Chen X, Lee H, Deshmukh SS, Yoganarasimha D, Savelli F, Knierim JJ. 2018. Egocentric coding of external items in the lateral entorhinal cortex. Science 362(6417):945-949.

Webb TJ, Rolls ET. 2014. Deformation-specific and deformation-invariant visual object recognition: pose vs identity recognition of people and deforming objects. Frontiers in Computational Neuroscience 8:37.

Whitlock JR. 2017. Posterior parietal cortex. Curr Biol 27(14):R691-R695.

Wirth S, Avsar E, Chiu CC, Sharma V, Smith AC, Brown E, Suzuki WA. 2009. Trial outcome and associative learning signals in the monkey hippocampus. Neuron 61(6):930-40.

Wirth S, Baraduc P. 2018. [Spatial orientation in the primate: I see, there I am]. Med Sci (Paris) 34(1):33-36.

Wirth S, Baraduc P, Plante A, Pinede S, Duhamel JR. 2017. Gaze-informed, task-situated representation of space in primate hippocampus during virtual navigation. PLoS Biol 15(2):e2001045.

Wiskott L, Sejnowski TJ. 2002. Slow feature analysis: unsupervised learning of invariances. Neural Comput 14(4):715-70.

Wyss R, Konig P, Verschure PF. 2006. A model of the ventral visual system based on temporal stability and local memory. PLoS Biol 4(5):e120.

Zhao D, Si B, Tang F. 2019. Unsupervised feature learning for visual place recognition in changing environments. Proceedings of the 2019 International Joint Conference on Neural Networks.

Zipser D, Andersen RA. 1988. A back-propagation programmed network that simulates response properties of a subset of posterior parietal neurons. Nature 331(6158):679-84.