Highlights

A Large-scale Neurocomputational Model of Spatial Cognition Integrating Memory with Vision

Micha Burkhardt, Julia Bergelt, Lorenz Gönner, Helge Ülo Dinkelbach, Frederik Beuth, Alex Schwarz, Andrej Bicanski, Neil Burgess, Fred H. Hamker

- Novel, systems-level approach integrates vision and spatial memory/imagery
- Integration of multiple brain areas gives rise to key aspects of spatial cognition
- Interfacing memory and vision through parietal areas improves object localisation
- Virtual environment opens up new options for the assessment of computational models

A Large-scale Neurocomputational Model of Spatial Cognition Integrating Memory with Vision

Micha Burkhardt[a,1,2], Julia Bergelt[a,1,2], Lorenz Gönner[b,c,1,2], Helge Ülo Dinkelbach[a,1,2], Frederik Beuth[a,1,2], Alex Schwarz[a,1,2], Andrej Bicanski[d,1,2], Neil Burgess[e,2], Fred H. Hamker[a,2,*]

[a] *Chemnitz University of Technology, 09107, Chemnitz, Germany*
[b] *Technische Universität Dresden, Faculty of Psychology, 01062, Dresden, Germany*
[c] *Technische Universität Dresden, Department of Psychiatry, 01307, Dresden, Germany*
[d] *Newcastle University, NE1 7RU, Newcastle upon Tyne, United Kingdom*
[e] *University College London, WC1E 6BT, London, United Kingdom*

Abstract

We introduce a large-scale neurocomputational model of spatial cognition called 'Spacecog', which integrates recent findings from mechanistic models of visual and spatial perception. As a high-level cognitive ability, spatial cognition requires the processing of behaviourally relevant features in complex environments and, importantly, the updating of this information during processes of eye and body movement. The Spacecog model achieves this by interfacing spatial memory and imagery with mechanisms of object localisation, saccade execution, and attention through coordinate transformations in parietal areas of the brain. We evaluate the model in a realistic virtual environment where our neurocognitive model steers an agent to perform complex visuospatial tasks. Our modelling approach opens up new possibilities in the assessment of neuropsychological data and human spatial cognition.

Keywords: brain-inspired neural networks, spatial reference transformation, parietal cortex, visual attention, spatial memory and imagery

[*] Corresponding author

[1] Equal contribution

[2] *E-mail addresses:*
micha.burkhardt@informatik.tu-chemnitz.de (Micha Burkhardt)
julia.bergelt@informatik.tu-chemnitz.de (Julia Bergelt)
lorenz.goenner@tu-dresden.de (Lorenz Gönner)
helge-uelo.dinkelbach@informatik.tu-chemnitz.de (Helge Ülo Dinkelbach)
frederik.beuth@informatik.tu-chemnitz.de (Frederik Beuth)
alex.schwarz@informatik.tu-chemnitz.de (Alex Schwarz)
andrej.bicanski@newcastle.ac.uk (Andrej Bicanski)
n.burgess@ucl.ac.uk (Neil Burgess)
fred.hamker@informatik.tu-chemnitz.de (Fred H. Hamker)

Preprint submitted to Elsevier *June 29, 2023*

1. Introduction

Uncovering the underlying mechanisms of spatial cognition involves a broad spectrum of research ranging from experimental studies with animals and humans to neurocomputational modelling. Spatial cognition in its broadest interpretation must solve various problems including object detection (Cavanagh, 2011), visual attention (Carrasco, 2011), eye- and body-movements (Land, 2009), as well as spatial memory and navigation (Burgess, 2008). This combination of tasks requires an intimate coupling between visual perception and cognition as outlined in the seminal work of Ballard et al. (1997), which suggests a variable binding of objects in the world to internal cognitive programs through deictic (”do-it-where-I'm-looking”) strategies. In visuospatial tasks, the issue of spatial reference frames also comes into play: While visual information is initially processed in a retinal reference frame, grasping often relies on body or limb centred reference frames (Pouget & Sejnowski, 1997), and navigation can even recruit world-centred (allocentric) reference frames (Avraamides & Kelly, 2008).

We propose that integrating mechanistic models into larger scale cognitive system models is required to explain such high-level cognitive functions. An example of this in a related domain is the Spaun architecture (Eliasmith et al., 2012), which implements a large-scale spiking network to output physical movements of a virtual robotic arm in a versatile set of cognitive tasks like digit recognition, serial working memory, or mental arithmetic. In the context of spatial cognition, previous bio-inspired models mostly focus on spatial navigation (Becker & Burgess, 2000; Byrne et al., 2007) and a few modelling approaches also exist in the field of bio-inspired robots, although with varying biological plausibility (Antonelli et al., 2014; Moulin-Frier et al., 2018). As overt behavior is typically the result of a coordinated activation involving many parts of the brain, attempts are required to not only integrate models, but also to improve the understanding of function across brain parts, which is limited when neurocomputational models are only studied in isolation.

A particular aspect we are interested in is the ability of humans to guide attention by long-term memory. Experimental studies have revealed that the hippocampus, via the parietal cortex, contributes to object detection (Summerfield et al., 2006; Salsano et al., 2021). However, most experimental studies and computational models that study attention and vision typically direct attention based on visual cues, but not based on long-term memory. From a conceptual point of view, the necessity for such an interaction of vision and memory has previously been outlined by Epstein et al. (2017), who argued that an effective use of a cognitive map requires to anchor such a map to the world. To the best of our knowledge, this interplay of memory and vision in a spatial context is yet to be explored by neurocomputational models.

In order to explore this interaction, we introduce the Spacecog model as a systems-level approach to spatial cognition and shed light on how multiple brain areas might interact with each other to display key elements of spatial memory and object detection. Built on the foundation of previous work done under the European research project ”Spatial Cognition” (Hamker, 2015), Spacecog builds upon three individual neurocomputational models: A model of attention including object recognition and object detection (Beuth, 2019), a model of perisaccadic space perception (Bergelt & Hamker, 2019), and a model of spatial memory and imagery (Bicanski & Burgess, 2018). Through this integration, we propose how parietal cortical areas could interface with visual and long-term memory areas to form a complex understanding of the surrounding environment. As visual information is initially encoded in a retinocentric reference frame the question arises how spatial memory, stored in a world-centered reference frame, can guide visual perception.

The individual parts of the Spacecog model are anatomically constrained and, as shown by the original publications, replicate experimental findings of neural mechanisms responsible for vision, attention, eye movements, and memory recall. By combining them into a large-scale neurocomputational model of spatial cognition, we aim to bridge the gap between previously disparate lines of research and particularly explore the putative role of the parietal cortex interfacing vision and memory. Acting as a case study for an increased understanding of the integration of brain areas, we propose how the brain deals with complex tasks in our everyday environment. We test the model on a functional level in a real-world like virtual environment, in which a neuro-cognitive agent has the task to successfully locate, encode, recall and re-localise objects in a realistic scene.

2. Methods

The neurocomputational model has been used to perform visuospatial computations for a neuro-cognitive agent (Figure 1a) which operates in a virtual environment (Figure 1b). We next introduce the virtual environment and explain the model and its functions.

2.1. Virtual Environment

The Unity game engine[3] was used to create the spatial environment (a child's room) and a cognitive agent, which we will refer to as Felice. Felice is connected to the neural network through a custom network interface built with Google's protocol buffers[4]. This extra step allows for the computational network to run on a separate Linux server, while Unity is running on a Windows computer, distributing the workload. Alternatively, it is possible to use virtualisation techniques (e.g. the Windows Subsystem for Linux, WSL2) to run the whole model on a single machine.

Pictured in Figure 1b;c, the main feature of the environment is a large desk with several toys placed on top, which can be recognised, remembered, and recalled by Felice. During simulations, Felice is externally instructed to walk into the vicinity of a random target object, which is placed among others on her desk. She first localises and encodes this object into memory, and is then instructed to walk to a different location. Once arrived, Felice will use object identity to recall information about the object location from memory. This subsequently allows Felice to walk back to the original position and to visually re-localise the object.

[3] https://unity.com/

[4] https://developers.google.com/protocol-buffers

By default, objects in the virtual environment are subject to a perspectival projection, which results from a 3D world being projected onto a 2D camera plane. This projection, however, results in artefacts of distorted proportions of objects, especially in the corners of an image. While the model can tolerate small deformations and still performs well in such cases, we mitigate any potential issues by introducing a spherical projection shader to more closely mimic human vision and to ensure position-invariant object proportions in the visual field (Figure 1c).

The visual, perceptual input from the virtual environment (as 408x308p colour images) is processed by the computational model, which returns commands for specific motor actions like positional changes (rotation and translation) or eye movements. Therefore, Felice can perceive objects from different angles and distances, as well as under different lighting. This creates a challenging, real-world-like environment for her to act in.

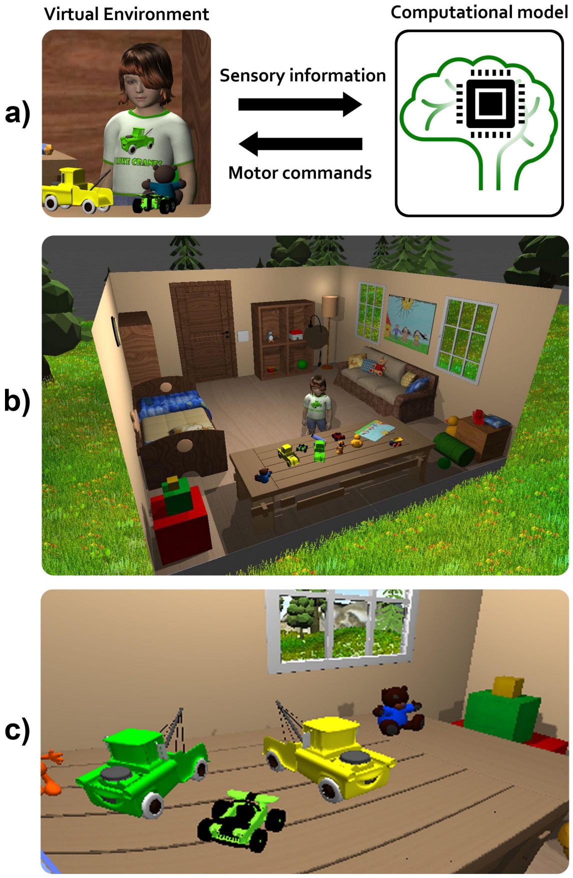

Figure 1: The virtual environment. a) The virtual environment provides sensory information which is sent to the model. The neural network then evaluates these data to form an internal, dynamic representation of the environment and outputs motor commands to be performed by the agent. b) The child's room of the cognitive agent called Felice. She is able to navigate and shift gaze as well as locate, remember, and recall multiple objects in this environment. c) Example of a visual image from the viewpoint of the agent. For simplicity, we do not model different resolutions of an object with respect to retinal eccentricity and also only use monocular vision (single eye camera).

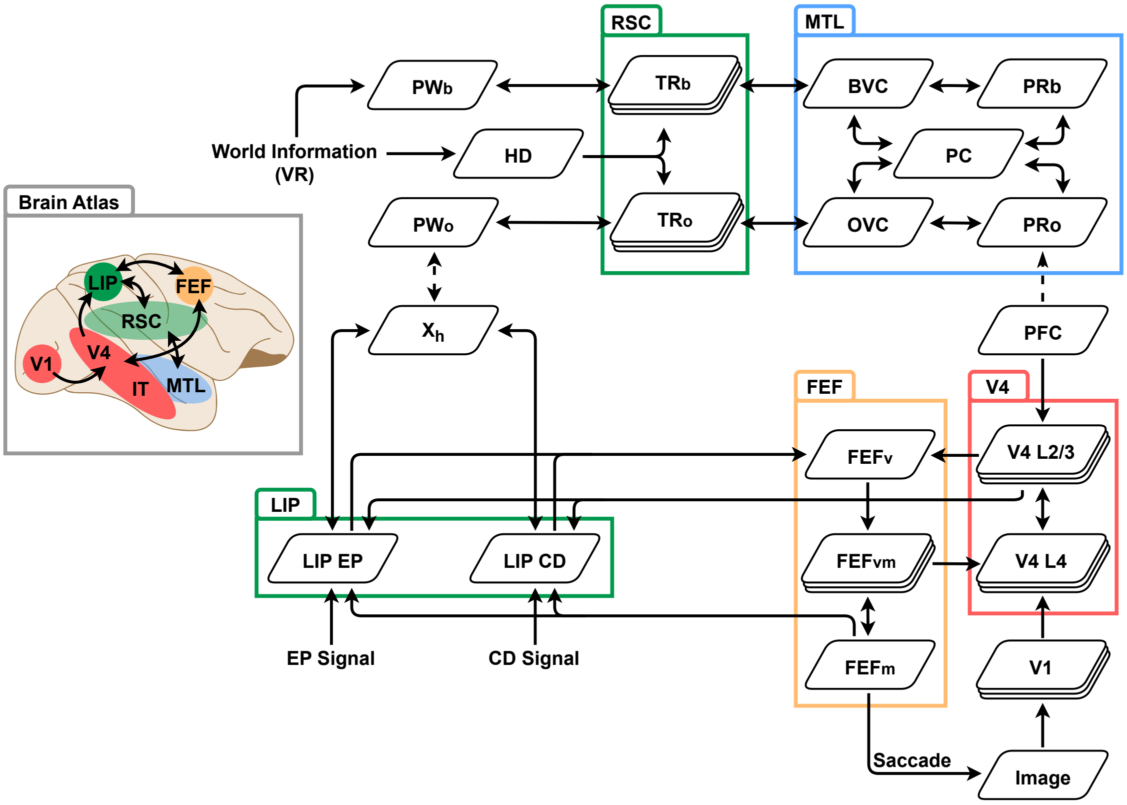

Figure 2: The Spacecog neurocomputational model. Different parts of the model strongly interact with each other based on anatomical constraints and functional purposes. The red and yellow parts of the model cover object detection and saccade planning, respectively. PFC provides a feature-based top-down bias to V4/IT neurons and information about object identity to PRo neurons, which are part of spatial long-term memory in MTL (blue). Attention emerges by the inherent reentrant dynamics in this system but is biased by the different top-down directed signals. Information from the saccade plan (CD) is sent to LIP (green) through X_{FEF}, where it is transformed into a head-centred reference frame with an eye position (EP) signal from the VR. Object location information is also transformed from an eye-centred to a head-centred reference frame in LIP and stored in a parietal priority map (X_h). This information is further passed to PW, where it is transformed into a world-centred reference frame via RSC (green) using head-direction. Object position, object identity, spatial boundaries, and the position of the agent in the room (place field activity) are encoded into memory in MTL. During memory recall, world-centred information from MTL is fed back to PWo via RSC and further into X_h, from where it acts as a spatial attention signal in V4/IT via LIP and FEF. If neural activity in the FEF movement (FEFm) cells exceeds a threshold, a saccade is triggered to the location indicated by the FEFm cells. The shift of the eyes is externally determined by saccade generator affecting the input image that visually samples the world. On the top-left, a lateral view on the brain areas is given. Not depicted is PW, which is postulated to be located in the precuneus. Brain image from Smith Breault (2020). Abbreviations: V1 - primary visual cortex, V4 - fourth visual cortex, IT - intraparietal cortex, PFC - prefrontal cortex, FEF - frontal eye fields (with visual, visuomovement, movement cell characteristics), LIP - lateral intraparietal cortex, EP - eye position, CD - corollary discharge, X_h - parietal priority map, PW - parietal window (objects, boundaries), HD - head direction cells, RSC - retrosplenial cortex, TR - transformation circuit (objects, boundaries), MTL - medial temporal lobe, BVC - boundary vector cells, OVC - object vector cells, PR - perirhinal neurons (objects, boundaries), PC - place cells. Solid arrows denote fully connected neural populations, while dotted arrows show connections which require additional (external) cues.

Spatial cognition is a large field of research. We here specifically aim to address the interplay of visuospatial and memory components. This requires an integration of object memory with visual perception including eye-centred visual processes, world-centred information of objects and space in long-term memory, as well as visual attention and object detection. For this, three individual models are integrated to form a large-scale neurocomputational model (Figure 2). These biologically rooted models were previously described in detail and have been extensively validated and compared with human and macaque experimental data. Even though some of these data are from different species, the model can be considered being a generic model of processes that are likely similar among different species.

Our integrated model can operate in two directions corresponding to processes of encoding and mental imagery: 1) In encoding processes, an object is searched by the agent via means of feature-based attention which alters the response profile of object cells in area V4/IT of the visual cortex. At the same time, this V4/IT information drives the frontal eye fields (FEF) for saccade target selection. Spatial information about this object is transformed from an eye-centred reference frame into a head-centred reference frame via the lateral intraparietal cortex (LIP) and, after being combined with environmental information in the parietal window (PW), transformed into a world-centred reference frame via the retrosplenial transformation circuit (RSC/TR). For long-term memory storage, this combined information of objects and space is encoded in an attractor network in the medial temporal lobe (MTL). 2) In processes of mental imagery, neural patterns from a previous encoding phase are re-instated in MTL by a cue-based memory retrieval using object identity. The retrieved patterns contain spatial information of the object and agent during encoding (their relative location to each other and their absolute locations relative to the environment). They are used for spatial navigation (here only the navigational goal) and furthermore transformed from world-centred into eye-centred reference frames via RSC and LIP for attentional control in FEF.

2.2.1. Object Recognition

To implement the capability of recognising and localising objects, the visual part of the computational model incorporates key elements of the ventral stream in the primate brain (Beuth, 2019). The input to the model is the agent's current visual field, which is a monocular RGB-image. Only daylight vision is considered, which in primates is represented by L,M, and S cones in the retina to process long (L), middle (M) and short (S) wavelengths of the visible light spectrum. For this purpose, the image is processed in area V1, which includes neurons organised in three channels. These channels are arranged in a retinotopic fashion and include cells for the red-green (L-M) and blue-yellow (LM-S) colour contrasts which are commonly found in the lateral geniculate nucleus (LGN) (Gegenfurtner & Kiper, 2003). In addition, the channels contain oriented edges which are derived from the image using Gabor filters (Jones & Palmer, 1987). Thus, they represent neurons with receptive fields commonly found in primary visual cortex (V1) simple cells (Jones & Palmer, 1987). The low level colour and orientation features within the field of view are then fed into higher visual areas for further processing (Figure 2, red parts).

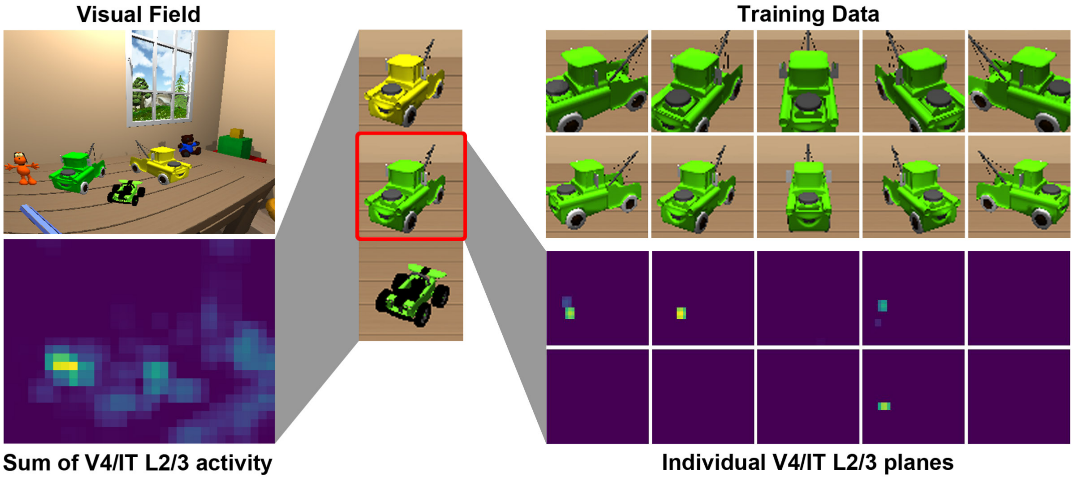

Figure 3: Object localisation for the green crane. Area V4/IT L2/3 consists of 30 neuronal layers/planes. Each plane encodes cells representing an object from a particular view-point (view-tuned cells). The number of planes is a result of the training procedure, as for each of the three objects weights were calculated for five different rotations and two sizes (here shown for the green crane). Thus, for each object we only use 10 images for training. The sum of neuronal activity in V4/IT L2/3 over all planes/objects (left) as well as activity in the individual planes for the green crane (right) are shown. When the agent is searching for the green crane (denoted by the red square), these planes are subject to feature-based attention from the prefrontal cortex. Their neural activity reflects the match of parts of the encoded object with the particular visual image, and the gain via feature-based attention.

Object recognition requires more narrowly tuned cells that respond selectively to an object or parts of it. Hence, V1 features need to be combined in higher visual areas by learning useful representations of objects. In our model, we simulate a higher visual area (V4/IT; Figure 2), representing high-level visual cortices such as the fourth visual cortex (V4) or the inferior temporal cortex (IT), with cells encoding object views (object-view tuned cells) as found in the inferior temporal cortex (Logothetis et al., 1995). V4/IT Layer 4 (V4/IT L4) encodes these object views, which are created by a convolution of the activities of V1 neurons with pre-learned weights. These weights were generated through a process called one-shot learning (Jamalian et al., 2016), which generated the weight matrix directly from the output of V1 complex cells in a prior learning phase. For training, only 10 images per object (five rotations, two sizes) were used to allow for some degree of invariant recognition. Furthermore, spatial pooling of these activities takes place in V4/IT Layer 2/3 (V4/IT L2/3). Like Layer 4, Layer 2/3 neurons have still a spatial organisation being selective for different parts of the image (Figure 3). Layer 2/3 neurons are subject to feature-based attention from the prefrontal cortex (PFC), enhancing the gain of neurons that respond to the target object.

This object recognition part is comparatively simple compared to deep neural networks. Thus, our emphasis is not on recognising a large number of objects, but to allow object recognition on a number of pre-selected objects with only a very small amount of training data (Figure 3). Technically, as further processing in the network is not dependent on the specific structure of the V1-V4/IT path, one could replace the image processing by a deep neural network to provide our model with a feedforward input into V4/IT L4, as we did not implement feedback connections back to V1.

2.2.2. Saccade Execution

Activities of object neurons in V4/IT L2/3 are pooled over objects to elicit spatially distributed neural activations in the frontal eye fields (FEF; Figure 2, yellow parts). Neural populations in FEF are responsible for the processing of spatial information and the preparation of eye-movements, particularly saccade target selection. This part is based on a model developed by Zirnsak et al. (2011).

Our model of the FEF is divided into three parts, namely FEF-visual (FEFv), FEF-visuomovement (FEFvm) and FEF-movement (FEFm), inspired by recordings from frontal eye-field neurons (Schall et al., 2004). From a functional perspective, FEFv indicates potentially relevant locations by taking the maximum activities over features in V4/IT L2/3. Feedforward soft-competition, combined with feedback from eye movement preparation in FEFm, activates FEFvm neurons. Feedforward projections from FEFvm to FEFm accompanied by strong lateral competitive interactions lead to the potential target of the upcoming saccade. If FEFm neurons increase their activation beyond a threshold, a saccade is executed towards the centre of gravity of the activation profile at that time. Given the saccade target, the actual movement of the eyes is then modelled by an extended version of the saccade generator of Van Wetter & Van Opstal (2008).

2.2.3. Attention

Among other tasks, our model is designed to perform object localisation supported by attentive dynamics, most notably feature-based and spatial attention. This model component is based on previous models that explain attention as an emergent result of neural dynamics, rather than postulating brain circuits that exclusively compute attention (Hamker, 2003, 2005b; Zirnsak et al., 2011) and is inspired by biased competition (Desimone & Duncan, 1995) and feature-similarity (Treue, 2001) frameworks of attention. The present model is built upon a microcircuit of attention proposed by Beuth & Hamker (2015), who compared and fitted their model to electrophysiological data of more than 10 different experiments that studied the mutual influence of stimuli placed within or near receptive fields in different states of attention. This is to date the most exhaustive comparison of a model with data recorded from neurons localised in different visual brain areas modulated by attention.

A top-down signal from PFC amplifies target-feature-specific activities in area V4/IT, independent of the location or size of features in the visual field and allows the selection of specific objects (Beuth, 2019). This can be seen in the sum of V4/IT L2/3 activity in Figure 3, where feature-based attention amplifies features of the green crane. In parallel, spatial attention emerges by feedback from FEFvm cells, which link spatial attention to the eye movement plan (Hamker et al., 2008) and can also account for attentional capture, based on the attention-related N2pc component of EEG recordings (Novin et al., 2021).

The here presented integrated model extends spatial attention with an additional loop between FEF and LIP. This extension by the LIP circuit is crucial for aspects of spatial cognition as LIP connects visual areas through parietal areas with MTL, where long-term information of objects and the environment are stored. If this information is recalled, LIP can, through coordinate transformations which are more closely described in the next section, generate an additional spatial attention signal to recalled locations of previously encountered and encoded objects. As this attentional signal does not require a feature-search in the entire visual field, it acts faster than spatial attention generated in the V4/IT-FEF loop. Also, this spatial attention signal is updated during eye movements to ensure that attention is directed to an object in space regardless of gaze position.

The general role of the LIP circuit in spatial attention has previously been motivated by Bisley & Goldberg (2003) and Goldberg et al. (2006). The computational model of spatial updating in the parietal cortex was first introduced by Ziesche & Hamker (2011) and later extended by Ziesche & Hamker (2014); Bergelt & Hamker (2016); Jamalian et al. (2017); Ziesche et al. (2017); Bergelt & Hamker (2019). The model has been compared to and motivated by studies exploring predictive remapping of attention (Rolfs et al., 2010), lingering of attention after saccades (Golomb et al., 2010), a combination of both (Jonikaitis et al., 2013), perisaccadic mislocalisation of briefly flashed stimuli (Van Wetter & Van Opstal, 2008), and saccadic suppression of displacement (Deubel et al., 1996).

2.2.4. Coordinate Transformation

Spatial tasks of embodied agents operating in, and interacting with the world require coordinate transformations between different reference frames. Generally speaking, we can distinguish between egocentric and allocentric (world-centred) reference frames. Egocentric reference frames include eye-centred or head-centred reference frames, while allocentric reference frames could relate to cardinal directions or visual landmarks. In our case, visual information about an object, initially processed in an eye-centred reference frame, could then be transformed into an allocentric reference frame for storage in long-term memory. It has been proposed that gain fields and radial basis functions (Figure 4) can perform these coordinate transformations between eye- and head-centred reference frames (Pouget & Sejnowski, 1997; Pouget et al., 2002), and diagonal connection patterns for this transformation have recently been observed in Drosophila (Lu et al., 2022). In our model, these coordinate transformations are performed in LIP (Ziesche & Hamker, 2011; Bergelt & Hamker, 2019) and RSC (Bicanski & Burgess, 2018) (Figure 2, green parts).

While the agent is searching for an object, retinal (eye-centred) input from area V4/IT is passed to LIP, along with a retinotopic spatial signal from FEF, a proprioceptive eye position (EP) signal encoding the eye position in a head-centred reference frame, and a corollary discharge (CD) signal encoding the eye displacement in an eye-centred reference frame. According to Ziesche & Hamker (2011) and Bergelt & Hamker (2019), the retinal signal from V4/IT L2/3 is fed into LIP maps, where it is gain-modulated by the CD signal (LIP CD) as well as the EP signal (LIP EP). This produces a combined representation of eye position and object position. Reading out the activity in LIP, we receive the perceived spatial position of an object in head-centred coordinates stored in $\mathrm{X_h}$ (Figure 4; $\mathrm{Stim_{head}}$). As mentioned above, this process can also be

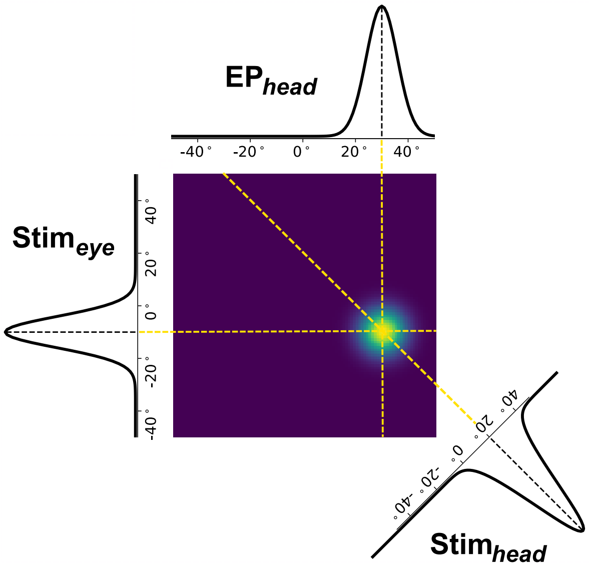

Figure 4: Coordinate transformation in a radial basis function network. In this example, a stimulus position in an eye-centred reference frame gets transformed into a head-centred reference frame. The eye fixates on a position at 30° (EP_{head}) and the eye-centred stimulus position is at -10° (Stim_{eye}). Thus, the resulting head-centred position of the stimulus is at 20° ($\text{Stim}_{\text{head}}$). Further, planned gaze shifts (and not only eye position) are also used for coordinate transformation. The same principle is also used for transformations between head- and world-centred reference frames. Importantly, this transformation can also be performed in the opposite direction (world-centred to head-centred or head-centred to eye-centred).

performed in a top-down fashion to transform a head-centred signal into an eye-centred signal, which is sent to FEFv to attend to the retinotopic position of a previously encoded object.

Object location information and the local environmental layout need to be combined with head direction to enable an unambiguous representation of objects and space. This is conducted in the spatial memory pathway of the model (Bicanski & Burgess, 2018), which demonstrates how neural representations of head-centred (egocentric) experiences interface with world-centred representations in long-term memory. The parietal areas of the brain, which we call 'parietal window' (PW), include head-centred representations of discrete objects (PWo) and boundaries (PWb). Here, objects refer to the three potential target objects and boundaries refer to the four walls of the room.

In addition to object information (PWo), during encoding/perception, the parietal window is also driven by high-level (head-centred) visual information of boundaries (PWb). This input is externally provided by the virtual environment and not explicitly modelled. The resulting activities in PW are fed into RSC, where the head direction signal (HD) provides gain-modulation to transform the egocentric representations into a world-centred reference frame, similar to related mechanisms proposed for the posterior parietal cortex (Pouget & Sejnowski, 1997; Whitlock et al., 2008). This circuit can also further be used in the opposite direction, which is required for processes of recall. World-centred information about boundaries and objects stored in MTL are then transformed back into a head-centred reference frame in PW via RSC.

2.2.5. Spatial Memory and Imagery

After visuospatial information is transformed into a world-centred reference frame through RSC, the resulting allocentric representations located in the medial temporal lobe (MTL; Figure 2, blue parts) can contribute to long-term memory (Bicanski & Burgess, 2018). Information in MTL consists of boundary information, which is encoded in so called boundary vector cells (BVCs) for the external boundaries of the environment (as an allocentric counterpart of PWb cells) and object vector cells (OVCs), which encode the position of objects and hippocampal place cells (PCs), which encode an allocentric position of the agent in space (see Bicanski & Burgess (2020) for a more in-depth review of the properties of vector coding cells in the brain). For a given spatial position encoded by PCs, an explicit subset of BVCs and OVCs are co-active to form a high-level representation of the spatial scene. By connecting co-active populations via Hebbian-like learning, an allocentric MTL attractor network is formed, which enables spatial long-term memory. Additionally, since BVCs and OVCs do not distinguish between specific boundaries and objects, perirhinal neurons (PR), high-level neurons of the ventral stream, code for the identity of boundaries (PRb) and objects (PRo). These allocentric representations can subsequently be used for memory recall. Cueing a previously encoded object in MTL enables the re-instated neuronal activities to drive the transformation circuit in the opposite direction, establishing egocentric representations to be reconstructed from memory and thus enabling spatial imagery through the parietal window. This egocentric information can then further be used as attentional input for LIP and FEF.

The basis of stable representations of self-location is the agent's perception of boundaries, which drives firing of BVCs. Their activity, in turn, activates corresponding place cell firing, in a manner consistent with empirical data (O'Keefe & Burgess, 1996) and with established computational models of place field generation by BVCs (Barry et al., 2006). This connectivity has been pre-trained dependent on the perceived layout of the room and the agent's location. Hence the agent treats walls as stable, while smaller objects can move. The configuration of BVCs that is consistent with the given location has synaptic connections with a cluster of place cells for that given location (and vice versa for recall). For the purposes of the integrated model, the spatial memory component assumes that parts of the visual system can extract the distance and egocentric directional information of boundaries from retinal inputs. This is not explicitly modelled and relies on cues from the VR (Fig. 2, world information). A network that performs these computations could be learned (Lian et al., 2023), but this mechanism is beyond the scope of the present manuscript.

2.2.6. Model Specification

Spacecog is built on the foundation of several previously published models, and detailed information about the neural models and their underlying assumptions can be found in these works as outlined in the previous sections. As these models have already been fitted to experimental data, most of the parameters in visual and spatial areas remain unchanged. The present model focuses on more holistic aspects, which allows us to explore more complex and complete tasks through the integration of memory and vision through parietal areas. This part of the model requires an interaction of visual and spatial neural populations, which operate in different coordinate systems. While head-centered information in both X_h and PWo is two-dimensional, the representation in PWo changes from a visual field (height and width) to a birds-eye spatial map (left/right, radial distance) as visible in Figure 6. Thus, while other populations in the model are fully connected on a neural level, this transition requires additional information. During bottom-up encoding processes, only the horizontal component of the X_h signal is used and supplemented with externally provided depth information from the VR before being used as an object cue for PWo. In return, this loses the height-dimension of the visual field, which is stored externally to be used during the back-transformation in the recall phase.

Further, a sensible balance between feature-based and spatial attentional mechanisms needs to be found. As feature-based attention originates in PFC and directly modulates V4/IT activities, the spatial attention pointer during recall originates in OVC and has to be looped back all the way into visual areas. For this, in addition to the already established recurrent V4/IT-FEF loop, we expanded the model by a recurrent V4/IT→LIP→FEFv→FEFvm→V4/IT loop, consistent with the idea that a connection between these areas could act as a simple representation of attentional priority, which is also fed back into visual areas (Bisley & Mirpour, 2019). Also the ventral stream model was adapted. Learning was conducted via the more simple and fast one-shot learning (Jamalian et al., 2016), rather than Hebb-type trace learning (Beuth, 2019). The pooling inside V1 and the pooling from V1→V4/IT was adapted, so the receptive fields of a V4/IT neuron are large enough to fit the relevant objects. Thus, some parameters in the mentioned areas were tuned by hand to facilitate such behaviour. As a result, the model is able to robustly encode objects into memory and to use this knowledge as spatial attentional information to enhance re-localisations of previously encoded objects. Further, we will also show that this structure of the model allows for the observation of perceptual neglect-like behaviour when a lesion is introduced in the parietal X_h priority map.

2.2.7. Model Implementation

The neurocomputational model is implemented with ANNarchy 4.7.1.1 (Vitay et al., 2015). ANNarchy (Artificial Neural Networks architect) is a neural simulator designed for distributed rate-coded or spiking neural networks. The user-interface is written in Python and uses an equation-oriented mathematical description of the neuron and synapse models. From this description, ANNarchy will generate efficient C++ code to perform the network simulation on parallel hardware.

We provide the complete source code for the model and the virtual environment, which is publicly accessible through `https://github.com/hamkerlab/Burkhardt2023_SpatialCognition`. With the provided code, the simulations introduced in the present paper can be replicated and freely modified (limited to the pre-trained spatial environment of the room and the three target objects). Due to the large size of the model, a complete description of the network can be found in the supplementary material. This includes the equations for all neural populations as well as all parameters and connections used. More detailed information about the neural models and their underlying assumptions can also be found in the previously published works (Bicanski &

Burgess (2018); Beuth (2019); Bergelt & Hamker (2019)).

3. Results

To evaluate the performance of our model, the cognitive agent Felice has to perform tasks in the virtual environment, which were developed to emulate a real-life situation requiring features of spatial cognition such as object localisation, attentive dynamics, coordinate transformations, and spatial memory and imagery. We first introduce the general structure of an integrated task combining these requirements, and later evaluate it through multiple experiments, modifying individual parts of our model and tasks.

3.1. General Task

Let us assume the following scenario: Felice first wants to play with one of her toys (e.g. a green toy crane), which she is able to localise among several other toys on her desk. She then gets distracted by another task and ends up at a different location in the room. From there, Felice decides she wants to again play with the toy crane, remembers where she initially found the toy and subsequently walks back to the location where she initially spotted the crane in order to localise it again. For this, we will use ego-, and allocentric information as outlined in the model description, but not relative information such as ”on the desk”. Decisions about ’when to do what’ are pre-defined, as decision making is not a particular focus of this study.

Generally, the task can be divided into an encoding phase and a recall phase. In the encoding phase, Felice has to find and encode an object into memory, which incorporates the entire process of using eye-centred visual information and transforming it into allocentric representation of objects and space in long-term memory. More specifically, starting from an arbitrary position, Felice walks into the vicinity of potential target objects. In this case, a random position within a circular area in front of her desk is assigned as a plausible position (Figure 5, middle). While Felice is able to move freely within the boundaries of her room, path planning and walking is not part of our model and was achieved through a simple A* search algorithm. Once she arrives in front of the desk, Felice aims to select one target among other potentially relevant objects. For simplicity, we here demonstrate this ability with three objects: A green and a yellow toy crane, as well as a green race car. Finding the target among a combination of these objects covers the main challenges for the object localisation, namely a similarity in shape and/or colour for the distractor objects. Additionally, features of the room itself can also be regarded as distractors, as neurons might respond to edges or colour gradients in them. After Felice successfully localises the object, its position and identity as well as the position of Felice are stored in long-term memory in MTL (by learning connections between place cells and object vector cells).

Figure 6 displays the structure and activities of neuronal populations for the encoding phase. First, the visual field is pre-processed in V1, features are extracted in V4/IT, and spatial attention emerges from the recurrent V4/IT-FEF loop. In parallel, this information is passed to LIP, where a coordinate transformation takes place to transform the object position from a retinotopic into a head-centred reference frame in $\mathrm{X_h}$. From there, object and spatial boundary information are formed in the parietal window and this (head-centred) information is then transformed into an allocentric reference frame via RSC and encoded into long-term memory in MTL populations.

Felice then walks to a different position, from where the recall phase shown in Figure 7 is triggered. The recall phase consists of Felice remembering, walking back to, and re-localising a previously encoded object. This process starts by applying an externally provided cue to the PRo neuron coding for the previously encoded object (”I want my green crane!”), which can be understood as an ”eyes-closed remembering”, with no interference from the current perceptual input. The resulting memory recall re-establishes information about the position of Felice and the memorised object from the time of encoding in the MTL attractor network. Specifically, Felice is then able to decode her previous position and body orientation from PC and HD activities and walks back to this recalled position. For this purpose, we read out PC and HD firing rates after memory recall and use this for external navigation. Once she arrives at the desk, Felice again searches for her green crane. The re-localisation of the target object can take place in three different ways: In the first way, Felice can use feature-based attention from PFC (as in the encoding phase shown in Figure 6) to re-localise the object. This can intuitively be construed as Felice remembering that she has previously seen the desired object from a particular position (e.g. close to the table), but she has no access to its exact location. In the second way, Felice uses recalled spatial memory information that is looped back through the PW (Figure 7) and being used for a spatial attention pointer generated in $\mathrm{X_h}$. The spatial attention pointer biases the neural dynamics within the visual system to guide visual search (without feature-based attention), corresponding to Felice recalling where the target object is located, without remembering its exact identity. Third, a combination of both options (feature-based attention combined with a spatial attention pointer) can be used, meaning Felice now recalls the location and relative direction as well as the identity of a previously encoded object.

To test these conditions, we introduce two different experimental settings (normal and cluttered scene), which cover the integration of vision and memory in slightly different scenarios. As the model operates in a realistic setting, we assess its behavioural performance through successful trial completions and the duration needed for each object localisation. The main purpose of the evaluation is to demonstrate the spatio-cognitive ability by means of a brain-inspired model. Further, we report and discuss differences observed in the sketched ways to perform the task.

3.2. Experiment 1: Spatial memory and object recognition in a normal scene

Experiment 1 is performed in an environment containing three potential target objects (Figure 8). The three possibilities of using attentional mechanisms for object localisation in the recall phase described above are used to asses the integration between vision and spatial memory.

3.2.1. Experiment 1.1: Spatial memory and object recognition using feature-based attention

Initially, Felice was asked to walk into the vicinity of potential target objects, where the first object localisation was performed randomly for one of the three objects. Thus, our model is run by setting an

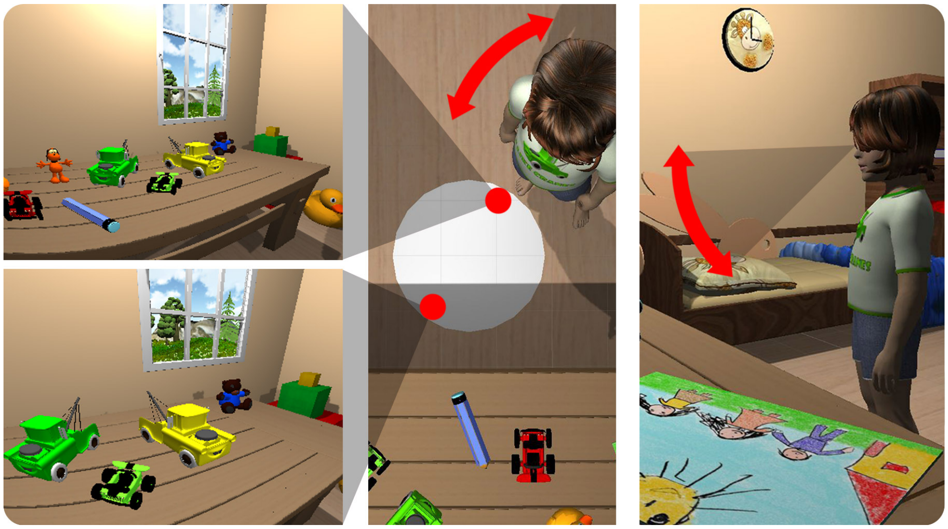

Figure 5: General scenario. The cognitive agent Felice performs a combined task of memory encoding and recall. Illustrated here is the encoding phase, in which Felice walks into the vicinity of target objects (random position within the white circle shown in the middle image). Depending on her position, body orientation and head tilt are adjusted to ensure the visibility of all potential target objects (middle, right). Variability in this adjustment results in different random views of the scene (left). The limiting factor for the positional variability is the spatial resolution of the visual layers V4/IT, and therefore the size of the objects in the visual field, which was controlled for by the allowed positions for object localisation (white circle).

activation for the target object in PFC, which allows feature-based attention to support target localisation. Then, after shifting gaze to the target object, the integrated model encodes the object and positional information into long-term memory. This corresponds to the encoding phase displayed in Figure 6 and was performed for total of 100 times. In each trial, a random target object and a random agent position (within the white circle displayed in Figure 5) were chosen, which resulted in a successful object localisation in 93% of trials (Table 1, first row).

In each trial, Felice then walked to a different position in the room, from which the encoded object was not visible and recalled her memory by activating the PRo cell of the previously encoded object. Part of this memory is her previous position (encoded by place cells) and the corresponding object location in the room (Figure 7). Felice then walked to the recalled position and performed a re-localisation of the previously encoded object. In this experiment, she used feature-based attention from PFC, and subsequently performed a saccade to the selected candidate object. If the object was correctly re-localised, the trial was labelled as successful. We therefore define a successful recall as a correct completion of the complete task, which includes the encoding and recall phase. This resulted in a success rate of 95% (Table 1, second row), which marginally differs from the encoding phase due to small deviations in Felice's position during recall.

We found the success in this experiment to be mainly dependent on two factors: First, the visual part of the model performing a correct object localisation and second, the spatial part of the model accurately encoding the positional information, enabling Felice to precisely return to the location of the previously encoded target object. However, as the second object localisation in this experiment also only required feature-based attention, the positional memory was not required to be extremely accurate. All errors in this experiment therefore were a result of the visual part of the model incorrectly localising the object. We will consider this result as a baseline and compare it to results from Experiment 1.2 and 1.3.

Table 1: Performance of the model (N=100).

Task	Experiment	Attention	Success rate	Simulation steps (M ± SD)
Encoding	1.1,1.2,1.3	Feature-based	93%	150 ± 57
Recall	1.1	Feature-based	95%	154 ± 59
Recall	1.2	Spatial	83%	172 ± 48
Recall	1.3	Spatial + feature-based	94%	125 ± 10

3.2.2. Experiment 1.2: Spatial memory and object recognition using a spatial attention pointer from memory

In the second experiment, rather than feature-based attention, a spatial attention pointer from the memory recall via LIP was used to perform the second object localisation after Felice returned to the previously encoded target object. During recall, allocentric information of the object position are transformed through RSC into head-centred activity in the parietal window. After Felice returns to the place where she encoded the target object, this information is subsequently used in X_h and LIP to generate a retinocentric spatial attention pointer. An advantage of using spatial attention provided by LIP is that it does not require extensive visual search to localise the target object.

We observed that without the aid of feature-based attention, the spatial attention pointer from long-term memory alone could generate a success rate of 83% (Table 1). Compared to Experiment 1.1, this is a slight decrease in performance, which was mostly caused by two factors: First, even small inaccuracies in the positional recall of Felice (decoded from PCs) were able to change the resulting visual field to a degree in which the spatial attention pointer was slightly misplaced. Second, inaccuracies in the recalled position due to a limited spatial resolution of neural populations (each PWo and OVC cell covers a 7° bin of the visual field) could also lead to a small, erroneous shift of the attention pointer, even though a weighted average approach was used to decode this information. Despite these limitations, only a few additional errors occurred, mostly in cases in which two objects were close to each other (Figure 8, Experiment 1.2).

3.2.3. Experiment 1.3: Spatial memory and object recognition using a combination of feature-based attention and a spatial attention pointer from memory

Experiment 3 combined both information sources about the object, namely spatial and feature-based attention, in the recall phase. This increased the performance back to the level of encoding and therefore mitigated errors previously introduced by spatial attention (Figure 8, Experiment 1.3). Additionally, as

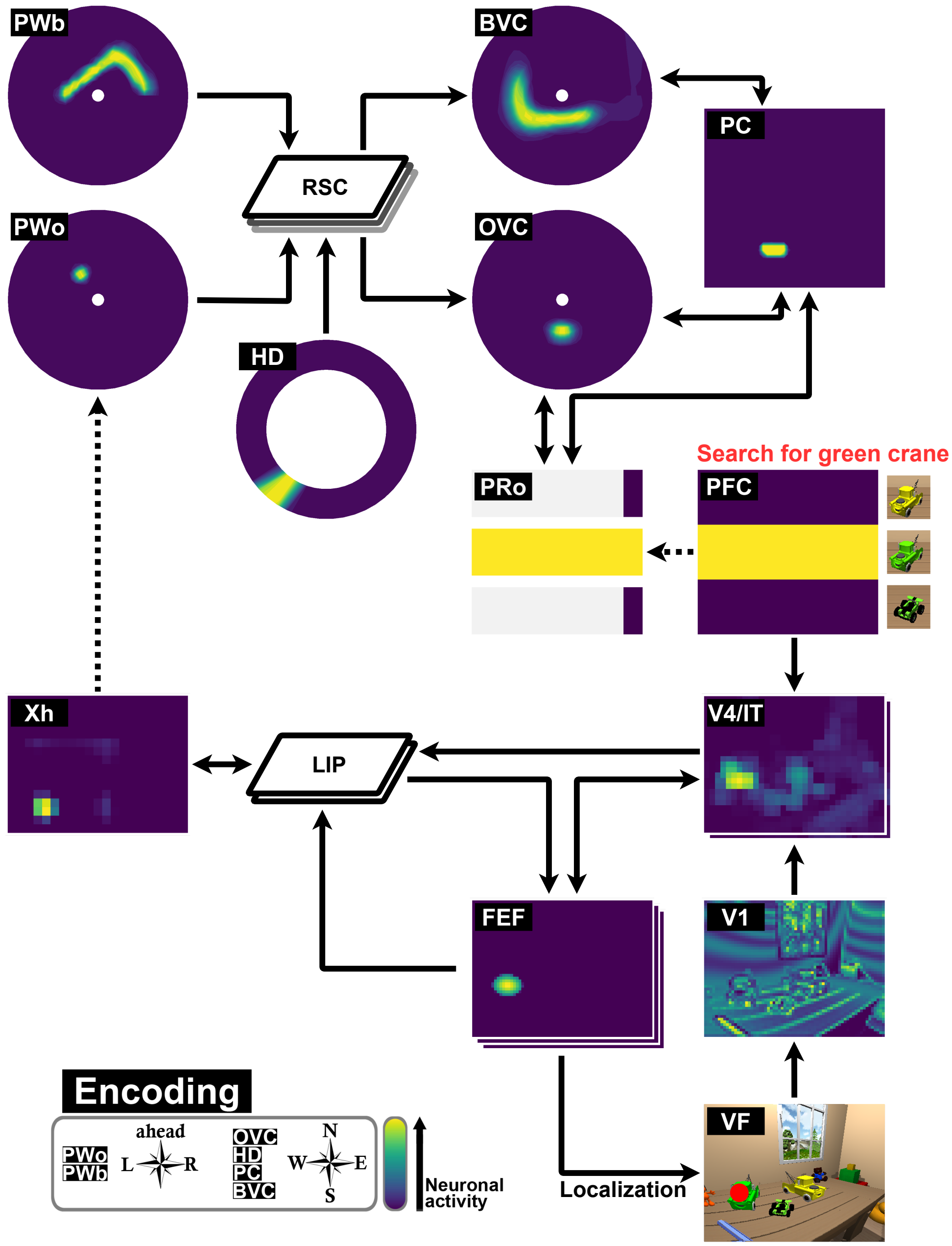

Figure 6: A representative subset of neural activity for the encoding phase of the general task of visual search and object memory (not included: V4/IT L4, FEFv, FEFvm, PRb). For encoding, the visual field (VF) is processed by V1 neurons and fed into higher visual areas. There, object neurons in V4/IT L2/3 are guided by feature-based attention from PFC and spatial attention emerges via FEFvm feedback to V4/IT until FEFm triggers a saccade towards the object. Spatial information is transformed from a retinocentric into a head centred reference frame in area LIP and gives rise to activity in X_h, where it is used as a head-centred input for parietal window object neurons (PWo) encoding the spatial position of the object relative to the agent (here ahead-left). This requires externally provided depth information from the virtual environment, while height information has to be saved for the recall phase. Combined with information of the boundaries of the room (PWb) as well as head direction (HD), this information is transformed through RSC, resulting in allocentric representations of the object location (OVC), boundaries (BVC), and agent position (PC) being established in MTL, where they are encoded into long-term memory.

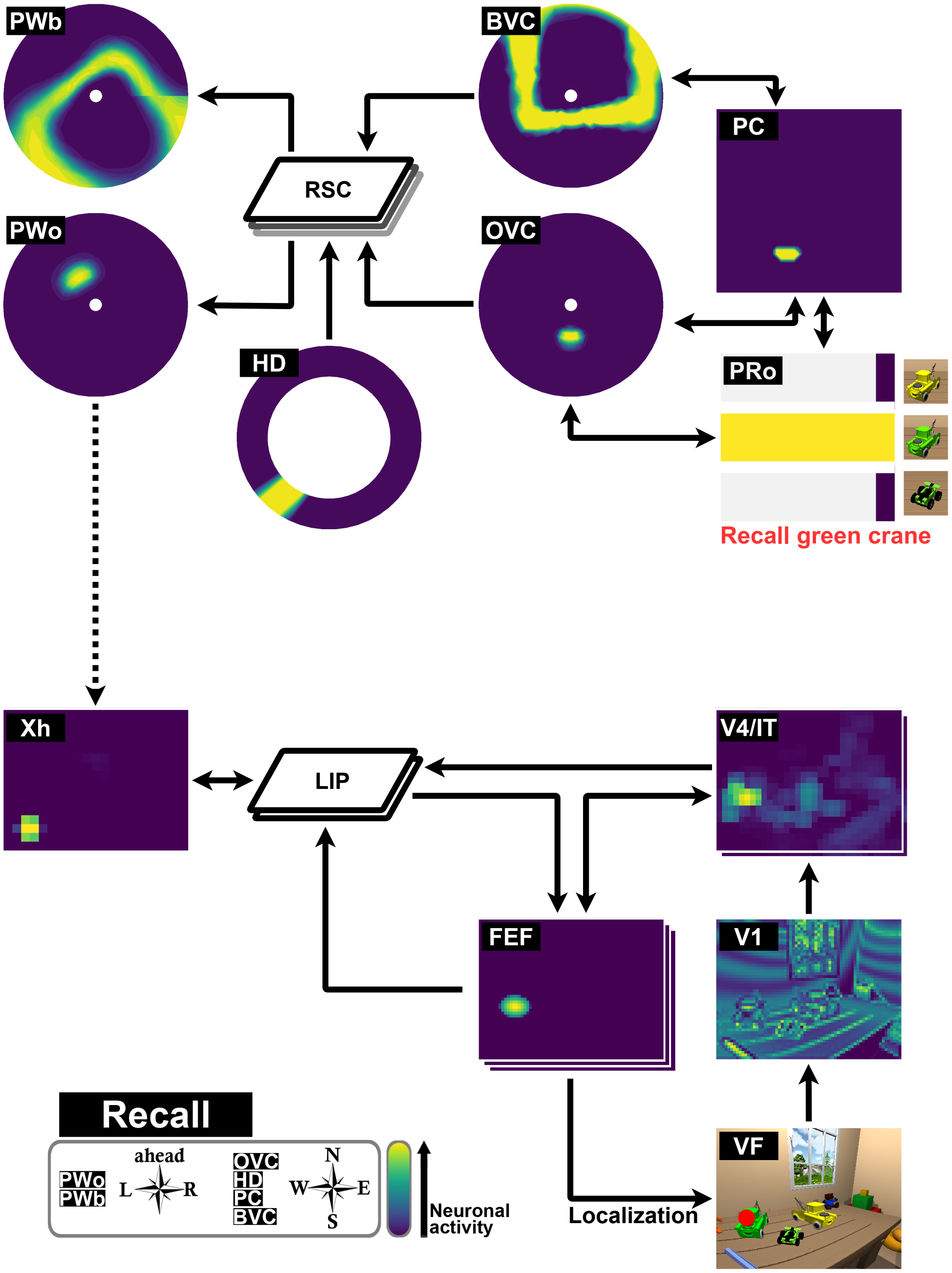

Figure 7: A representative subset of neural activity for the recall phase of the general task of visual search and object memory (not included: V4/IT L4, FEFv, FEFvm, PRb). From a remote position in the room, a recall of the previously encoded object is triggered through PRo activation. This re-establishes activity in the MTL attractor network (OVC, BVC, PC) as well as in HD populations. Through RSC, this information is transformed into a head-centred reference frame in PWo and PWb. Information of the object position can then be fed back into X_h, from which a retinocentric spatial attention pointer can be established in area LIP. With this top-down attention signal being applied on FEFv, the position of the object can then be decoded from the build-up movement neurons in FEFm. Once an activity threshold in FEFm is reached, a saccade is performed towards the target object.

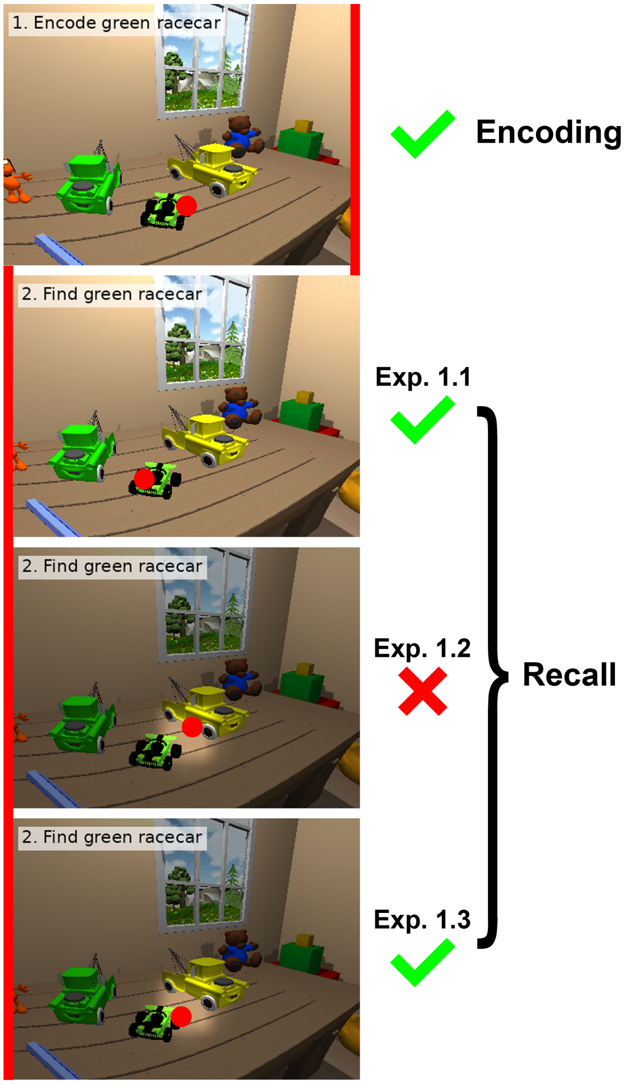

Figure 8: Experiment 1. Felice initially encodes the green race car. Small inaccuracies in the spatial memory can lead to a slightly different agent position during recall and therefore also a slight shift in the visual field (the red bars display the alteration in the visual field between encoding and recall phase, which result in a shift to the right). Results from the recall phase of all three experiments are shown (Experiment 1.1 uses only feature-based attention, Experiment 1.2 uses only a spatial attention pointer, and Experiment 1.3 utilises both feature-based and spatial attention). Shifts in the visual field do not affect feature-based attention used in Experiment 1.1, but can lead to ambiguous situations in Experiment 1.2, in which the spatial attention pointer is placed between two objects. If the spatial attention pointer is however combined with feature-based attention (Experiment 1.3), these ambiguities can be resolved.

the spatial attention pointer was quickly available to assist during feature search, the time required for a successful re-localisation was reduced by 27% (Table 1). Therefore, the spatial attention pointer was able to guide the object localisation effectively, while feature-based attention compensated potential ambiguities through inaccuracies introduced through memory and imagery processes.

3.2.4. Neural dynamics for target selection

As the observed time for target selection varies in Experiments 1.1, 1.2, and 1.3 (Table 1), with the combination of feature-based and spatial attention being fastest, we analysed the temporal dynamics in the model (Figure 9). Feature-based attention operates across the entire visual scene and increases the gain of those neural responses where visual input and target template matches. Thus, neural activity in V4/IT is enhanced at the target location. Spatial attention recalled from memory traverses via $\mathrm{X_h}$ and LIP into the visual system and increases the activation at the target location if the recall from memory is correct. Recurrent dynamics lead to an exchange of activity across the whole visual parts, but they converge in FEF which enforces saccade target selection from FEFv, FEFvm to FEFm cells. If feature-based attention is used (Experiment 1.1 and 1.3), V4/IT has a higher activity than in Experiment 1.2, where only a spatial attention pointer is used for the localisation of the target object. In contrast, a spatial attention pointer leads to higher input from LIP in Experiment 1.2 and 1.3. Together, this results in the fastest rise of activity in FEFv for Experiment 1.3 and therefore the earliest initialisation of a saccade through FEFm among all three experiments.

3.3. Experiment 2: Spatial memory and object recognition in a cluttered scene

The previous experiments were all performed in the same general scene setting, where only the three target objects were presented in both encoding and recall phase as shown in Figure 8. Experiment 2 changes this by cluttering the scene between the memory encoding and the recall phase by placing additional toys onto the desk hiding the targets (Figure 10). This results in a much more challenging scenario and thus allows us to gain further insight into the interaction of memory and vision. Again, we distinguish between the three possibilities using attentional mechanisms for the object localisation in the recall phase. The summary of the results for this experiment is shown in Table 2.

3.3.1. Experiment 2.1: Feature-based attention

As the encoding phase is identical to previous experiments, we again observe a success rate of 93%. However, in the recall phase the object localisation now has to be performed in the cluttered scene, which dropped the performance to 27%, while the time required for a successful localisation increased to 231 steps, combined with a significant higher standard deviation, when only feature-based attention was used. As no spatial memory was used to aid the object localisation, the reduction in performance can solely be attributed to the visual model being unable to rely on enough features of the target objects, which are now substantially covered by other objects.

3.3.2. Experiment 2.2: Spatial attention from memory

In comparison to Experiment 1.2, this study uses only a spatial attention pointer from memory to perform the object localisation in the recall phase. This results in a further drop in performance to 14% and an increased localisation time of 239 steps. In addition to features of the target objects being covered by the

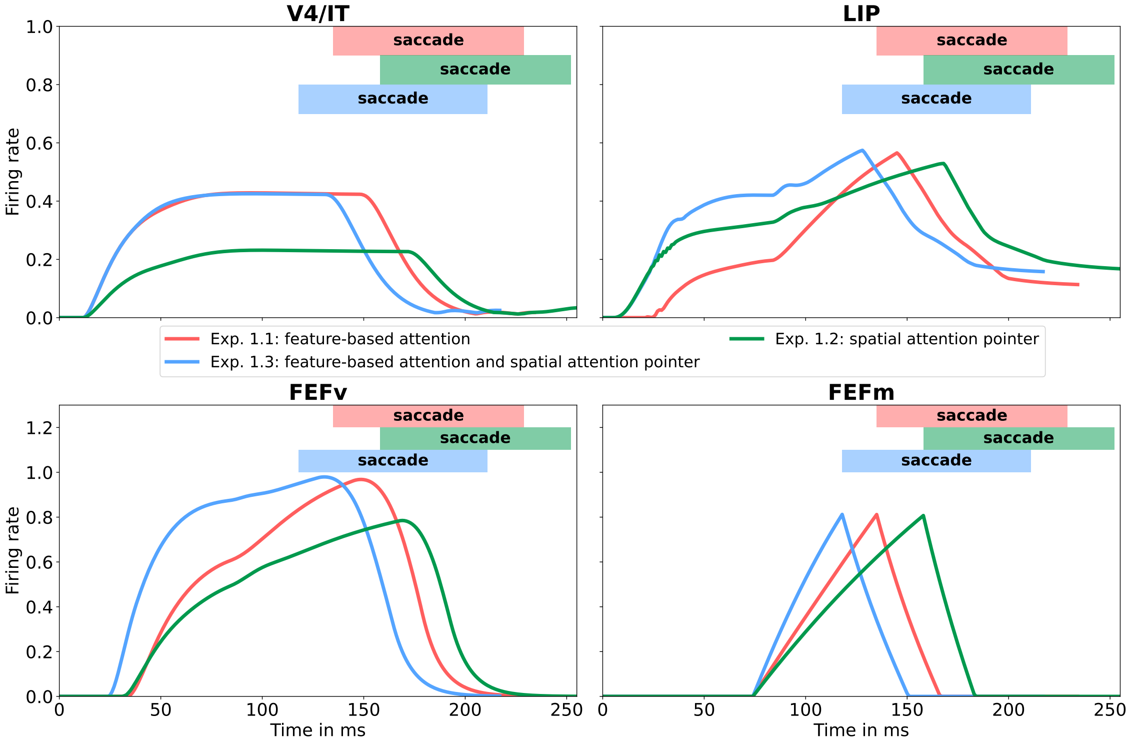

Figure 9: Analysis of temporal dynamics that lead to different reaction times across experiments. Plotted traces reflect the activation at the target location in different parts of the model. In conditions where visual search recruits feature-based attention, V4/IT activity increases due to the match of the visual input with the target template. In conditions where a spatial attention pointer from memory is recalled, the activation of LIP cells at the target location is increased. As FEFv cells collect information from those different parts of the model and pass it to FEFm to enforce a final decision about the saccade target, they reflect both biases in their activation. A saccade can be initialised fastest, if both feature-based attention and a spatial attention pointer are used. Shown are feature independent, pooled firing rates of V4/IT, Layer 2/3 (top left), and firing rates of LIP (top right) representing the target location, which both serve as input to FEFv (bottom left), as well as the firing rates of FEFm (bottom right), which trigger the saccade. Activation of a typical trial in each of the three Experiments 1.1 (red), 1.2 (green), 1.3 (blue) are plotted over time. The period of the saccade is marked for each experiment.

distractor objects, small inaccuracies in the position of the spatial attention pointer further contribute to the reduction in performance like in Experiment 1.2. It is to note that, while previous unsuccessful trials of object localisation almost always meant the selection of an incorrect object, in this experiment 84% of errors result from a timeout (FEFm activity not reaching a saccade threshold after 600 simulation steps). This is a direct limitation introduced by the small codebook of the visual model, which only includes three pre-trained objects, and therefore has no knowledge about the identity of the additional distractor objects. As V4/IT feeds the FEF, not much activity is transmitted across this pathway, which leads to an overall lesser activity in the visual system including FEFm. An additional putative method of compensation would be stronger self-excitation parameters in the saccade system to enforce saccade targets even into weakly activated areas. However, we kept all parameters unchanged to directly compare the different experiments.

3.3.3. Experiment 2.3: Combination of feature-based and spatial attention from memory

This simulation combines the attentional mechanisms of feature-based attention and the spatial attention pointer from memory during recall in a cluttered scene. The performance in this simulation increases to 48% while the average time required for a successful object localisation is reduced to 194 simulation steps. The main novelty of this simulation can be seen in comparison to Experiment 1, where the combination of both attentional mechanisms in Experiment 1.3 only recovered the reduction in performance back to the baseline level. This increase in performance for Experiment 2.3 further underscores the advantages of integrating spatial memory with vision, especially in challenging environments.

Table 2: Performance in cluttered scenes (N=100).

Task	Experiment	Attention	Success rate	Simulation steps (M ± SD)
Encoding	2.1,2.2,2.3	Feature-based	93%	150 ± 57
Recall	2.1	Feature-based	28%	231 ± 73
Recall	2.2	Spatial	14%	239 ± 99
Recall	2.3	Spatial + feature-based	48%	194 ± 69

3.4. Experiment 3: Visual Neglect

This simulation aims to further highlight the biological plausibility of our model by demonstrating that a simple impairment in a parietal area leads to similar behaviour as observed in patients with visual neglect.

Visual neglect is one of the most notable impairments resulting from damage to parietal areas of the brain, and is known to cause impairments in directional processes of attention and localisation of objects, resulting in a lack of responses to stimuli in parts of the visual field (Bartolomeo, 2007). This is thought to correspond to damage in a parietal priority map, which integrates goal and stimulus related signals for spatial selection (Bays et al., 2010). In our model, we can observe similar effects by introducing an impairment to X_h neurons corresponding to the left half of the visual field. This is implemented by removing all connections between LIP and the left half of X_h neurons, generating visual perceptual neglect.

Figure 11 depicts a simulation with two identical objects. Since feature-based attention favours both objects, the model converges on one of the cranes, depending on the exact spatial arrangement (position of agent and objects). In the depicted scenario, the left crane is chosen as the preferred stimulus, which is mostly visible in FEFm and X_h population activity (Figure 11, top). Consequently, a saccade is made to the left crane (visualised by the red dot in Figure 11, top right)

Introducing left visual neglect in an identical simulation results in only the right crane being active in the X_h priority map, which then leads to a reduced response of the left crane in LIP. Subsequently, through the recurrent LIP-FEF loop, this also creates a bias in FEF (Figure 11, bottom). There, bottom-up visual

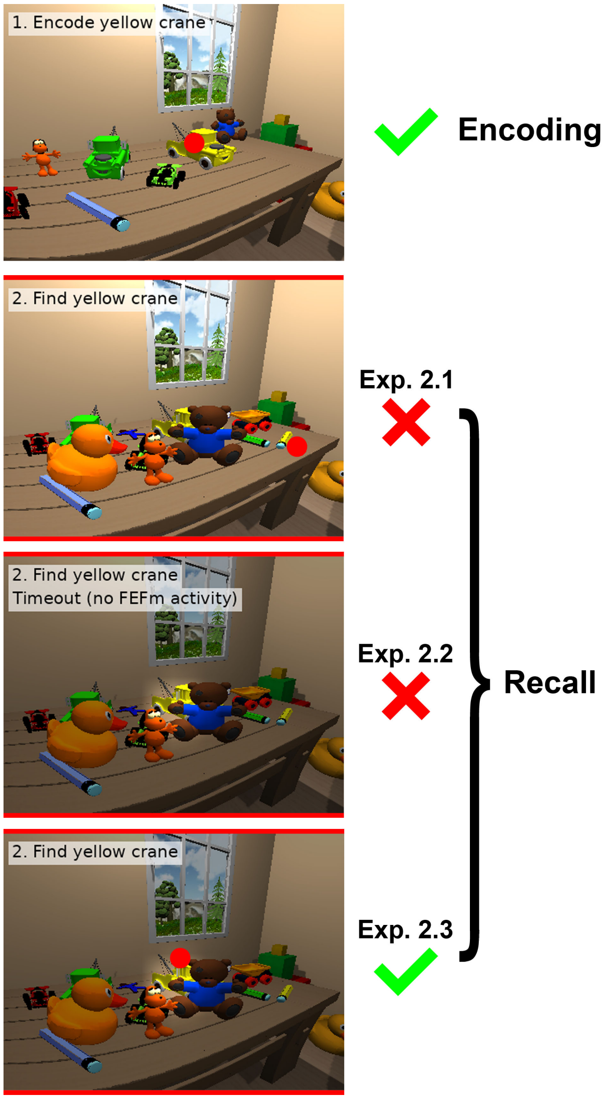

Figure 10: Experiment 2. In this condition, the scene is cluttered in the recall phase with additional distractor objects which cover the original target objects to a large degree. This creates a highly challenging task. In Experiment 2.1, the model is not able to re-localise the yellow crane, as most of its features are hidden behind the teddy bear. Experiment 2.2 with only spatial attention also fails despite a correct spatial attention pointer as no FEFm activity is formed due to a lack of V4/IT activity. Only in experiment 2.3 a successful re-localisation of the yellow crane is performed due to the combined application of feature-based and spatial attention. The red bars indicate the alteration in the visual field during recall, which in this case is the result of Felice recalling a position slightly closer to the table compared to the encoding phase. Here, this deviation is negligible and does not result in a misplacement of the spatial attention pointer.

input is modulated by attentional input from LIP, which leads to a saccade directed to the crane on the right (red dot in Figure 11, bottom right). Thus, although object recognition and initial spatial attention via FEF are unaffected by the impairment, a bias emerges in the additional recurrent LIP-FEF loop, which results in behaviour similar to visual neglect. Additionally, as X_h forms a bridge between visual and spatial areas, only the position of the right object will be passed to the parietal window and is encoded into memory. Visual neglect is therefore also present in memory and imagery.

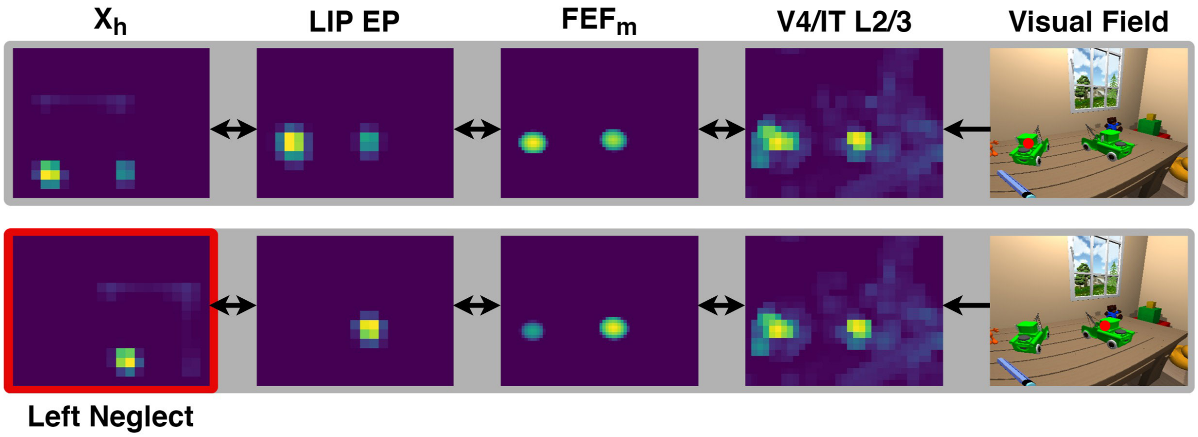

Figure 11: Visual neglect. In the top row, a typical simulation is shown, which results in the left crane being selected for a potential saccade (FEFm). When left side visual neglect is introduced in X_h, activities for the right object are projected back into FEF via LIP, while reentrant processing between LIP and FEF is weakened in the left visual field (bottom row), and as a result, the right object is selected as the saccade target.

4. Discussion

We have introduced the Spacecog model, a biologically motivated, large-scale neurocomputational model of spatial cognition, which we tested and evaluated in a real-world-like virtual environment. Via a coherent processing stream incorporating perceptual vision processes, attentive dynamics, and spatial memory and imagery, Spacecog is able to display key traits of spatial cognition. The underlying individual models were previously verified on their own and are motivated and grounded by anatomical, behavioural, and physiological data. While aspects specific to these individual components have already been described in previous publications, we here focused on the integration and interplay between memory and vision through parietal areas.

In three experiments, interactions between visual and spatial areas were evaluated and it was shown how the integration enables the agent to successfully perform tasks of object localisation and imagery. In all experiments, the agent was able to robustly detect and memorise objects. The introduction of a spatial attention pointer from memory by itself was able to generate a high success-rate during recall, but also introduced an increase in the time required for localisation due to the interplay of model components (Experiment 1.2). An integrated use of spatial and feature-based attention combined the advantages of a quick availability of the spatial attention signal from memory with the accuracy of feature-based attention to allow for a faster and robust re-localisation of objects (Experiment 1.3). Additional advantages of integrating spatial memory with vision were further explored in a cluttered environment, which showed that this integration is crucial for an adequate performance in even more challenging tasks (Experiment 2).

Notably, in conditions with only feature-based attention (Experiment 1.1/2.1), the agent was already successful in localising the target object. It therefore is important to clarify that the conditions for search with only feature-based attention were chosen to be optimal, as it was ensured that the agent was in close proximity to the objects, and the target object was ensured to be in her field of view. Without this, under normal conditions, a more extensive visual search including overt orienting responses would have been necessary. This also implies that during phases of recall, the recalled position and head direction are always used to guide the agent back to the target object, even when no direct interface between memory and vision through PW and LIP was established (Experiment 1.1/2.1). The good performance in only feature-based attention was also based on the fact that we did not use heavily cluttered scenes, and feature-based attention was still effective in guiding attention to the target. If we were to use more difficult search scenes, the search process would require multiple saccades, and thus the benefit of spatial memory would be more obvious. In such cases, the model would require an additional circuit to implement inhibition of return (Hamker, 2005a).

However, the main novelty is not the use of spatial or feature-based signals for object recognition, but the ability to establish spatial and object memory and to use this memory to guide vision. If agents are able to recall and use information about spatial proximity, gaze direction, feature-based attention, and spatial attention, they can accurately and efficiently re-localise previously encoded objects. This is a combination which has not yet been demonstrated in previous biologically motivated models. We underlined the robustness of this ability by allowing variability in the encoding process, resulting in different views of the scene. Furthermore, consistent with the idea that neglect results from damage in a parietal priority map (Bays et al., 2010), it was also shown that parietal lesions in our model can produce neglect-like behaviour (Experiment 3). Our integrated model therefore extends the mechanisms by which previous models accounted for spatial representational neglect (Byrne et al., 2007; Bicanski & Burgess, 2018) to neglect in the visual field.

The present model underlines the importance of the parietal cortex as an interface between vision and memory. Early concepts of the parietal cortex have already emphasised its role in providing 'where' information about the object (Mishkin et al., 1983), but were later extended with respect to actions towards objects (Milner & Goodale, 1995) and visual attention (Colby & Goldberg, 1999; Gottlieb, 2007), and more recently also with its role in episodic memory retrieval (Becker & Burgess, 2000; Cabeza et al., 2008; Sestieri et al., 2017; Connor & Knierim, 2017).

Further, the present model can be compared with experiments on human spatial cognition that show that long term spatial memory interacts with visual attention behaviourally, and reflects parieto-prefrontal activity related to attention and hippocampal and parahippocampal activity related to the benefit from long term memory (Summerfield et al., 2006). Our model can address the behavioural advantage for memory-cued locations (Experiment 1.3 and 2.3) and also its relation to the activity in different regions. A prediction from our model might be that hippocampal and retrosplenial activity correlate with performance more strongly when the target was previously seen in location from a different viewpoint, so that purely visual memories cannot give an advantage. Additionally, behavioural hypotheses could further be tested in identical virtual environments, as gaze behaviour in the context of locomotion was recently shown to be highly similar in virtual environments and the real world (Drewes et al., 2021).

Even though the complexity of the parietal and temporal cortex is much beyond what we cover with our model, our account proposes a framework of how memory recall can directly guide visual perception by means of transformation from allocentric to egocentric reference frames and visual attention. Thus, our Spacecog model demonstrates for the first time an integrated account of memory and vision.

Despite the model already covering some aspects of spatial cognition, it is by no means complete. Felice only uses monocular vision and we do not extract any depth information from the image by stereo vision or optic flow. Thus, despite her ability to recall and visit her recalled position in space, her understanding of space is limited due to missing depth information. Hence, in the present model, boundary information (the walls of the room) is supplemented by the VR and not the result of visual perception. The Spacecog model on which she operates also does have only a limited form of scene memory, while humans can store a large number of scene representations in visual long-term memory (Konkle et al., 2010), which may support self-localisation and allow to locate objects as part of the scene context (Hollingworth, 2007).

A further limiting factor of the model is computational complexity, which is most visible in the interaction of spatial and visual areas. Due to limits in the spatial accuracy of neural populations, small inaccuracies can occur during encoding and recall, which can lead to shifts in the position of the agent and its visual field, or in the placement of spatial attention. Increasing neural populations to more realistic numbers would be an obvious solution in this regard, however a performance-accuracy trade-off has to be made. Further, the development of biologically plausible but still efficient methods of object recognition is still an active research domain (Teichmann et al., 2021), while machine learning methods that rely on supervised learning are presently more powerful.

With respect to attention, it has been shown that the best possible search template are not necessarily the features of the to be searched object, but those features which best discriminate the target from distractor objects (Navalpakkam & Itti, 2007; Maith et al., 2021). Our present version of this model does not make use of learning a suitable top-down feature-based attention signal, given the context of a scene.

Finally, an important area of present and future research is the general flexibility of the agent. As outlined in the general task, the structure of the task is fairly fixed and Felice by herself is not given the ability to decide about her goals, plans and the outcomes of her actions. Thus, our model does not include reinforcement learning or other means of action selection. Although Felice is performing well in the specified task, this relies mostly on externally provided cues of where to initially walk and which object to attend to. Desired objects in encoding and recall are externally cued in the corresponding neural population and no intrinsic goal-directed behaviour is shown, as the required cognitive structures are not currently part of the model.

4.1. Conclusion

The combination of visual, attentional and spatial components successfully bridges a gap between previously disparate areas of neurocomputational modelling. By introducing parietal areas as an interface between spatial and visual areas, this most notably creates the novelty of memory guided visual attention, which at this level has so far only been addressed by the presented model. In addition to the questions explored above, spatial information of previously encoded objects can now be used to explore attentional processes across eye movements. This can open up many new avenues concerning the interpretation of neuropsychological data in complex tasks of spatial memory and attention. The integrated model therefore also provides a unified framework for visuospatial tasks and can further be used as a powerful tool for the assessment of a broad spectrum of biologically rooted hypotheses concerning human spatial cognition.

5. Data Availability

Code and data are available under `https://github.com/hamkerlab/Burkhardt2023_SpatialCognition`. Due to the size of the network, we will provide the model description as an ANNarchy report in the supplementary information.

6. Declaration of Interest

The authors declare no competing interests.

7. Acknowledgements

A first version of this model has been funded by the European Union's Seventh Framework Programme, grant number FP7-ICT 600785 Spatial Cognition. Funding by the German Research Foundation (DFG, 416228727) - SFB 1410 Hybrid Societies and by the Saxony State Ministry of Science and Art (SMWK3-7304/35/3-2021/4819) research initiative "Instant Teaming between Humans and Production Systems" on the basis of the budget passed by the deputies of the Saxony parliament, allowed us to further develop and verify the model. Additional funding was granted to Neil Burgess by the ERC advanced grant NEUROMEM.

8. Author Contributions

Micha Burkhardt: Conceptualisation, Methodology, Software, Validation, Formal Analysis, Data Curation, Writing - Original Draft, Writing - Review & Editing, Visualisation. **Julia Bergelt:** Conceptualisation, Methodology, Software, Validation, Writing - Review & Editing. **Lorenz Gönner:** Conceptualisation, Methodology, Software, Validation, Writing - Review & Editing. **Helge Ülo Dinkelbach:** Conceptualisation, Methodology, Software, Validation, Writing - Review & Editing. **Frederik Beuth:** Conceptualisation, Methodology, Validation, Writing - Review & Editing. **Alex Schwarz:** Methodology, Software, Validation.

Andrej Bicanski: Conceptualisation, Methodology, Writing - Review & Editing. **Neil Burgess:** Writing - Review & Editing, Resources, Supervision, Funding acquisition. **Fred H. Hamker:** Conceptualisation, Writing - Original Draft, Writing - Review & Editing, Resources, Supervision, Project administration, Funding acquisition

References

Antonelli, M., Gibaldi, A., Beuth, F., Duran, A. J., Canessa, A., Chessa, M., Solari, F., Del Pobil, A. P., Hamker, F. H., Chinellato, E., & Sabatini, S. P. (2014). A hierarchical system for a distributed representation of the peripersonal space of a humanoid robot. *IEEE Transactions on Autonomous Mental Development*, *6*(4), 259-273. doi: https://doi.org/10.1109/TAMD.2014.2332875

Avraamides, M. N., & Kelly, J. W. (2008). Multiple systems of spatial memory and action. *Cognitive Processing*, *9*(2), 93-106. doi: https://doi.org/10.1007/s10339-007-0188-5

Ballard, D. H., Hayhoe, M. M., Pook, P. K., & Rao, R. P. N. (1997). Deictic codes for the embodiment of cognition. *Behavioral and Brain Sciences*, *20*(4), 723–742. doi: https://doi.org/10.1017/S0140525X97001611

Barry, C., Lever, C., Hayman, R., Hartley, T., Burton, S., O'Keefe, J., Jeffery, K., & Burgess, N. (2006). The boundary vector cell model of place cell firing and spatial memory. *Reviews in the Neurosciences*, *17*(1-2), 71–98. doi: https://doi.org/REVNEURO.2006.17.1-2.71

Bartolomeo, P. (2007). Visual neglect. *Current Opinion in Neurology*, *20*(4), 381-386. doi: https://doi.org/10.1097/WCO.0b013e32816aa3a3

Bays, P. M., Singh-Curry, V., Gorgoraptis, N., Driver, J., & Husain, M. (2010). Integration of Goal- and Stimulus-Related Visual Signals Revealed by Damage to Human Parietal Cortex. *Journal of Neuroscience*, *30*(17), 5968-5978. doi: https://doi.org/10.1523/JNEUROSCI.0997-10.2010

Becker, S., & Burgess, N. (2000). Modelling spatial recall, mental imagery and neglect. In T. Leen, T. Dietterich, & V. Tresp (Eds.), *Advances in neural information processing systems* (Vol. 13). MIT Press.

Bergelt, J., & Hamker, F. H. (2016). Suppression of displacement detection in the presence and absence of eye movements: A neuro-computational perspective. *Biological Cybernetics*, *110*(1), 81-89. doi: https://doi.org/10.1007/s00422-015-0677-z

Bergelt, J., & Hamker, F. H. (2019). Spatial updating of attention across eye movements: A neurocomputational approach. *Journal of Vision*, *19*(7). doi: https://doi.org/10.1167/19.7.10

Beuth, F. (2019). *Visual attention in primates and for machines - neuronal mechanisms* (Doctoral dissertation, Department of Computer Science. Technische Universität Chemnitz). Retrieved from `https://nbn-resolving.org/urn:nbn:de:bsz:ch1-qucosa2-356553`

Beuth, F., & Hamker, F. H. (2015). A mechanistic cortical microcircuit of attention for amplification, normalization and suppression. *Vision Research*, *116*, 241-257. doi: https://doi.org/10.1016/j.visres.2015.04.004

Bicanski, A., & Burgess, N. (2018). A neural-level model of spatial memory and imagery. *eLife*, *7*, e33752. doi: https://doi.org/10.7554/eLife.33752

Bicanski, A., & Burgess, N. (2020). Neuronal vector coding in spatial cognition. *Nature reviews. Neuroscience*, *21*(9), 453-470. doi: https://doi.org/10.1038/s41583-020-0336-9

Bisley, J. W., & Goldberg, M. E. (2003). Neuronal activity in the lateral intraparietal area and spatial attention. *Science*, *299*(5603), 81-86. doi: https://doi.org/10.1126/science.1077395

Bisley, J. W., & Mirpour, K. (2019). The neural instantiation of a priority map. *Current Opinion in Psychology*, *29*, 108-112. doi: https://doi.org/10.1016/j.copsyc.2019.01.002

Burgess, N. (2008). Spatial cognition and the brain. *Annals of the New York Academy of Sciences*, *1124*, 77-97. doi: https://doi.org/10.1196/annals.1440.002

Byrne, P., Becker, S., & Burgess, N. (2007). Remembering the past and imagining the future: A neural model of spatial memory and imagery. *Psychological Review*, *114*(2), 340-375. doi: https://doi.org/10.1037/0033-295X.114.2.340

Cabeza, R., Ciaramelli, E., Olson, I. R., & Moscovitch, M. (2008). The parietal cortex and episodic memory: an attentional account. *Nature reviews. Neuroscience*, *9*(8), 613-625. doi: https://doi.org/10.1038/nrn2459

Carrasco, M. (2011). Visual attention: The past 25 years. *Vision Research*, *51*(13), 1484-1525. (Vision Research 50th Anniversary Issue: Part 2) doi: https://doi.org/10.1016/j.visres.2011.04.012

Cavanagh, P. (2011). Visual cognition. *Vision Research*, *51*(13), 1538-1551. (Vision Research 50th Anniversary Issue: Part 2) doi: https://doi.org/10.1016/j.visres.2011.01.015

Colby, C., & Goldberg, M. (1999). Space and attention in parietal cortex. *Annual Review of Neuroscience*, *22*, 319-349. doi: https://doi.org/10.1146/annurev.neuro.22.1.319

Connor, C. E., & Knierim, J. J. (2017). Integration of objects and space in perception and memory. *Nature Neuroscience*, *20*(11), 1493–1503. doi: https://doi.org/10.1038/nn.4657

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, *18*, 193-222. doi: https://doi.org/10.1146/annurev.ne.18.030195.001205

Deubel, H., Schneider, W. X., & Bridgeman, B. (1996). Postsaccadic target blanking prevents saccadic suppression of image displacement. *Vision Research*, *36*(7), 985-996. doi: https://doi.org/10.1016/0042-6989(95)00203-0

Drewes, J., Feder, S., & Einhäuser, W. (2021). Gaze during locomotion in virtual reality and the real world. *Frontiers in Neuroscience*, *15*, 656913. doi: https://doi.org/10.3389/fnins.2021.656913

Eliasmith, C., Stewart, T. C., Choo, X., Bekolay, T., Dewolf, T., Tang, Y., & Rasmussen, D. (2012). A Large-Scale Model of the Functioning Brain. *Science*, *338*, 1202-1205. doi: https://doi.org/10.1126/science.1225266

Epstein, R. A., Patai, E. Z., Julian, J. B., & Spiers, H. J. (2017). The cognitive map in humans: spatial navigation and beyond. *Nature Neuroscience*, *20*, 1504-1513. doi: https://doi.org/10.1038/nn.4656

Gegenfurtner, K. R., & Kiper, D. C. (2003). Color vision. *Annual Review of Neuroscience*, *26*(1), 181-206. doi: https://doi.org/10.1146/annurev.neuro.26.041002.131116

Goldberg, M. E., Bisley, J. W., Powell, K. D., & Gottlieb, J. (2006). Chapter 10 saccades, salience and attention: the role of the lateral intraparietal area in visual behavior. In S. Martinez-Conde, S. Macknik, L. Martinez, J.-M. Alonso, & P. Tse (Eds.), *Visual perception* (Vol. 155, p. 157-175). Elsevier. doi: https://doi.org/10.1016/S0079-6123(06)55010-1

Golomb, J. D., Pulido, V. Z., Albrecht, A. R., Chun, M. M., & Mazer, J. A. (2010). Robustness of the retinotopic attentional trace after eye movements. *Journal of Vision*, *10*(3). doi: https://doi.org/10.1167/10.3.19

Gottlieb, J. (2007). From thought to action: The parietal cortex as a bridge between perception, action, and cognition. *Neuron*, *53*(1), 9-16. doi: https://doi.org/10.1016/j.neuron.2006.12.009

Hamker, F. H. (2003). The reentry hypothesis: linking eye movements to visual perception. *Journal of Vision*, *3*(14), 808-816. doi: https://doi.org/10.1167/3.11.14

Hamker, F. H. (2005a). The emergence of attention by population-based inference and its role in distributed processing and cognitive control of vision. *Computer Vision and Image Understanding*, *100*(1), 64-106. (Special Issue on Attention and Performance in Computer Vision) doi: https://doi.org/10.1016/j.cviu.2004.09.005

Hamker, F. H. (2005b). The reentry hypothesis: The putative interaction of the frontal eye field, ventrolateral prefrontal cortex, and areas v4, it for attention and eye movement. *Cerebral Cortex*, *15*(4), 431-447. doi: https://doi.org/10.1093/cercor/bhh146

Hamker, F. H. (2015). Spatial Cognition of Humans and Brain-inspired Artificial Agents. *KI - Künstliche Intelligenz*, *29*, 83-88. doi: https://doi.org/10.1007/s13218-014-0338-8

Hamker, F. H., Zirnsak, M., Calow, D., & Lappe, M. (2008). The peri-saccadic perception of objects and space. *PLOS Computational Biology*, *4*(2). doi: https://doi.org/10.1371/journal.pcbi.0040031

Hollingworth, A. (2007). Object-position binding in visual memory for natural scenes and object arrays. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(1), 31. doi: https://doi.org/10.1037/0096-1523.33.1.31

Jamalian, A., Bergelt, J., Dinkelbach, H. U., & Hamker, F. H. (2017). Spatial attention improves object localization: A biologically plausible neuro-computational model for use in virtual reality. *2017 IEEE International Conference on Computer Vision Workshops*. doi: https://doi.org/10.1109/iccvw.2017.320

Jamalian, A., Beuth, F., & Hamker, F. H. (2016). The performance of a biologically plausible model of visual attention to localize objects in a virtual reality. In *International Conference on Artificial Neural Networks - ICANN 2016, Lecture Notes in Computer Science 9887* (pp. 447–454). doi: https://doi.org/10.1007/978-3-319-44781-0_53

Jones, J. P., & Palmer, L. A. (1987). An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex. *Journal of Neurophysiology*, *58*(6), 1233-1258. doi: https://doi.org/10.1152/jn.1987.58.6.1233

Jonikaitis, D., Szinte, M., Rolfs, M., & Cavanagh, P. (2013). Allocation of attention across saccades. *Journal of Neurophysiology*, *109*(5), 1425-34. doi: https://doi.org/10.1152/jn.00656.2012

Konkle, T., Brady, T. F., Alvarez, G. A., & Oliva, A. (2010). Scene memory is more detailed than you think: The role of categories in visual long-term memory. *Psychological Science*, *21*(11), 1551-1556. doi: https://doi.org/10.1177/0956797610385359

Land, M. F. (2009). Vision, eye movements, and natural behavior. *Visual neuroscience*, *26*(1), 51-62. doi: https://doi.org/10.1017/s0952523808080899

Lian, Y., Williams, S., Alexander, A. S., Hasselmo, M. E., & Burkitt, A. N. (2023). Learning the vector coding of egocentric boundary cells from visual data. *Journal of Neuroscience*. doi: https://doi.org/10.1523/JNEUROSCI.1071-22.2023

Logothetis, N. K., Pauls, J., & Poggio, T. (1995). Shape representation in the inferior temporal cortex of monkeys. *Current Biology*, *5*(5), 552-563. doi: https://doi.org/10.1016/s0960-9822(95)00108-4

Lu, J., Behbahani, A. H., Hamburg, L., Westeinde, E. A., Dawson, P. M.,Lyu, C., Maimon, G., Dickinson, M. H., Druckmann, S., & Wilson, R. I. (2022). Transforming representations of movement from body- to world-centric space. *Nature*, *601*, 98–104. doi: https://doi.org/10.1038/s41586-021-04191-x

Maith, O., Schwarz, A., & Hamker, F. H. (2021). Optimal attention tuning in a neuro-computational model of the visual cortex–basal ganglia–prefrontal cortex loop. *Neural Networks*, *142*, 534–547. doi: https://doi.org/10.1016/j.neunet.2021.07.008

Milner, D., & Goodale, M. (1995). *The Visual Brain in Action*. Oxford University Press. doi: https://doi.org/10.1093/acprof:oso/9780198524724.001.0001

Mishkin, M., Ungerleider, L. G., & Macko, K. A. (1983). Object vision and spatial vision: two cortical pathways. *Trends in Neurosciences*, *6*, 414-417. doi: https://doi.org/10.1016/0166-2236(83)90190-X

Moulin-Frier, C., Fischer, T., Petit, M., Pointeau, G., Puigbo, J., Pattacini, U., Low, S. C., Camilleri, D., Nguyen, P., Hoffmann, M., Chang, H. J., Zambelli, M., Mealier, A., Damianou, A., Metta, G., Prescott, T. J., Demiris, Y.,Dominey, P. F., & Verschure, P. F. M. J. (2018). Dac-h3: A proactive robot cognitive architecture to acquire and express knowledge about the world and the self. *IEEE Transactions on Cognitive and Developmental Systems*, *10*(4), 1005-1022. doi: https://doi.org/10.1109/TCDS.2017.2754143

Navalpakkam, V., & Itti, L. (2007). Search goal tunes visual features optimally. *Neuron*, *53*(4), 605–617. doi: https://doi.org/10.1016/j.neuron.2007.01.018

Novin, S., Fallah, A., Rashidi, S., Beuth, F., & Hamker, F. H. (2021). A neuro-computational model of visual attention with multiple attentional control sets. *Vision Research*, *189*, 104-118. doi: https://doi.org/10.1016/j.visres.2021.08.009

O'Keefe, J., & Burgess, N. (1996). Geometric determinants of the place fields of hippocampal neurons. *Nature*, *381*, 425-428. doi: https://doi.org/10.1038/381425a0

Pouget, A., Deneve, S., & Duhamel, J. R. (2002). A computational perspective on the neural basis of multisensory spatial representations. *Nature Reviews Neuroscience*, *3*(9), 741-747. doi: https://doi.org/10.1038/nrn914

Pouget, A., & Sejnowski, T. J. (1997). Spatial Transformations in the Parietal Cortex Using Basis Functions. *Journal of Cognitive Neuroscience*, *9*(2), 222-237. doi: https://doi.org/10.1162/jocn.1997.9.2.222

Rolfs, M., Jonikaitis, D., Deubel, H., & Cavanagh, P. (2010). Predictive remapping of attention across eye movements. *Nature Neuroscience*, *14*(2), 252-259. doi: https://doi.org/10.1038/nn.2711

Salsano, I., Santangelo, V., & Macaluso, E. (2021). The lateral intraparietal sulcus takes viewpoint changes into account during memory-guided attention in natural scenes. *Brain Structure and Function*, *226*, 989–1006. doi: https://doi.org/10.1007/s00429-021-02221-y

Schall, J., Thompson, K., Bichot, N., Murthy, A., & Sato, T. (2004). Visual processing in the macaque frontal eye field. In C. E. C. Jon H. Kaas (Ed.), *The primate visual system* (p. 205-230). CRC Press. doi: https://doi.org/10.1201/9780203507599

Sestieri, C., Shulman, G. L., & Corbetta, M. (2017). The contribution of the human posterior parietal cortex to episodic memory. *Nature reviews. Neuroscience*, *18*(3), 183-192. doi: https://doi.org/10.1038/nrn.2017.6

Smith Breault, M. (2020). Monkey brain. *Zenodo*. Retrieved from https://doi.org/10.5281/zenodo.3926117

Summerfield, J. J., Lepsien, J., Gitelman, D. R., Mesulam, M. M., & Nobre, A. C. (2006). Orienting attention based on long-term memory experience. *Neuron*, *49*(6), 905-916. doi: https://doi.org/10.1016/j.neuron.2006.01.021

Teichmann, M., Larisch, R., & Hamker, F. H. (2021). Performance of biologically grounded models of the early visual system on standard object recognition tasks. *Neural Networks*, *144*, 210-228. doi: https://doi.org/10.1016/j.neunet.2021.08.009

Treue, S. (2001). Neural correlates of attention in primate visual cortex. *Trends in Neurosciences*, *24*(5), 295-300. doi: https://doi.org/10.1016/S0166-2236(00)01814-2

Van Wetter, S. M., & Van Opstal, A. J. (2008). Experimental test of visuomotor updating models that explain perisaccadic mislocalization. *Journal of Vision*, *8*(14). doi: https://doi.org/10.1167/8.14.8

Vitay, J., Dinkelbach, H., & Hamker, F. H. (2015). Annarchy: a code generation approach to neural simulations on parallel hardware. *Frontiers in Neuroinformatics*, *9*, 19. Retrieved from https://doi.org/10.5281/zenodo.6417924 doi: https://doi.org/10.3389/fninf.2015.00019

Whitlock, J. R., Sutherland, R. J., Witter, M. P., Moser, M.-B., & Moser, E. I. (2008). Navigating from hippocampus to parietal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(39), 14755-14762. doi: https://doi.org/10.1073/pnas.0804216105

Ziesche, A., Bergelt, J., Deubel, H., & Hamker, F. H. (2017). Pre- and post-saccadic stimulus timing in saccadic suppression of displacement – a computational model. *Vision Research*, *138*. doi: https://doi.org/10.1016/j.visres.2017.06.007

Ziesche, A., & Hamker, F. H. (2011). A computational model for the influence of corollary discharge and proprioception on the perisaccadic mislocalization of briefly presented stimuli in complete darkness. *Journal of Neuroscience*, *31*(48), 17392-17405. doi: https://doi.org/10.1523/JNEUROSCI.3407-11.2011

Ziesche, A., & Hamker, F. H. (2014). Brain circuits underlying visual stability across eye movements - converging evidence for a neuro-computational model of area lip. *Frontiers in Computational Neuroscience*, *8*, 25. doi: https://doi.org/10.3389/fncom.2014.00025

Zirnsak, M., Beuth, F., & Hamker, F. H. (2011). Split of spatial attention as predicted by a systems-level model of visual attention. *European Journal of Neuroscience*, *33*(11), 2035-2045. doi: https://doi.org/10.1111/j.1460-9568.2011.07718.x