University of Trento
CIMeC – Centre for Mind/Brain Sciences

Doctoral School in Cognitive and Brain Sciences
(XXXVII Cycle)

What's in a cognitive map? Action representations in the hippocampal-entorhinal system

Alexander Eperon

Supervisors: Christian Doeller, Stephanie Theves and Roberto Bottini

Acknowledgements

This thesis exists thanks to my supervisors. Christian, thank you for welcoming me into your lab and trusting me with some big, very magnetic, machines. Stephanie, thank you for many great discussions and meticulous feedback. Roberto, thank you for fostering my enthusiasm and guiding it into practicable experiments. I look forward to more years of incomprehensible sketches while we try to figure out what our minds are up to.

Thank you to all the lab animals for their support, in particular Simone and Yangwen. Rebekka, for battling through fMRI analyses together. Giulia, for ready coffee and chats. Giorgia and Flavio, stay afloat. Fede (Dr Trains) and Giuli, time for a drink. Paco, shh.

This thesis owes much to the original CIMeC fellowship, Andrea, Chiara, Ludo and Nicola, for all the late nights telling stories and asking odd questions. This is the bedrock of science.

These four years would not have been the same without many more dear people, near and far. You know who you are. Special thanks go to Lea and Steffen, for selfless love and support; to Valerio and Giulia, for the years of escapism; to Nick and Ari, for the magic of lemons and limes, and to Mara, for shaping who I am.

Finally, thank you above all to Grandma, Mum and Dad, Giles, Robin, Felicity, and Ed. In a family of academics, thank you for teaching me to be kind.

Abstract

Spatially selective cell types in the mammalian entorhinal cortex are often taken as building blocks of a 'cognitive map', a mental representation of relational information. Nonetheless, relating these representations to behaviour, especially in non-spatial domains, has proven difficult. In this thesis, it will be argued that entorhinal representations of task structure integrate action information in support of flexible behaviour. Moreover, these representations appear connected to gaze behaviour and rotate across task contexts.

In a first study, I used functional magnetic resonance imaging (fMRI) to test for an entorhinal representation of learned actions in a conceptual space. To this end, I taught participants to associate numerically labelled states with specific mathematical operations, and examined neural pattern similarity as a function of action distribution similarity. Entorhinal neural pattern similarities scaled with action similarity. Moreover, these representations were tied to gaze behaviour, suggesting a potential link between the use of eye movements to navigate conceptual spaces and structural representations in the entorhinal cortex.

To understand these results better, a second study was carried out in which animal images were associated with context-dependent eye movement patterns in two different contexts. Firstly, I replicated the previous finding that the entorhinal cortex represents possible actions, showing that the entorhinal cortex encodes action similarities. Secondly, I tested action across task contexts and found evidence that neural representations rotated across graph contexts.

The present work, therefore, demonstrates that representations of relational structure in the entorhinal cortex are tied to action in non-spatial and visuospatial domains. The identified action representations are linked to gaze behaviour and appear to rotate in response to changes in task context.

Contents

Introduction

Mammals get around. Successful navigation is essential to ensuring survival, and the mammalian hippocampus and entorhinal cortex appear to employ a toolkit of specialised cell types to facilitate this (Bellmund et al., 2018). However, the role of these brain regions appears to stretch beyond spatial navigation to other domains, encompassing shifts in conceptual or visual spaces (Behrens et al., 2018; Bellmund et al., 2018; Bottini and Doeller, 2020). Moreover, the entorhinal cortex seems relatively robust to perturbation through environmental or stimulus changes (Fyhn et al., 2008; Baram et al., 2021) and has been linked to associative learning (Klingberg et al., 1994). In light of this, it has been proposed that the entorhinal cortex encodes the relational structure of our experiences for re-use across different stimulus sets (Stachenfeld et al., 2017; Whittington et al., 2020). Nonetheless, there is as yet little human experimental evidence connecting entorhinal representations to future actions in visual or conceptual spaces.

As rodents move around a flat arena, neurons in the medial entorhinal cortex release action potentials in a remarkable hexagonal pattern that tiles space at regular intervals (Hafting et al., 2005). Other cells in the medial temporal lobes and surrounding regions fire in response to individual locations, heading direction (Taube et al., 1990) or vector distances from objects (Bicanski & Burgess, 2020). The degree of specificity is striking: in the subiculum, for instance, neurons encode not just corners, but specific features such as angle and wall height (Sun et al., 2024). Seen together, these cells look like building blocks for a complete representation of allocentric space (Bakermans et al., 2025). Recent work has suggested that the entorhinal cortex also represents distance and direction information in non-spatial domains (Bellmund et al., 2018). Based on this, it has been suggested that the hippocampus and entorhinal cortex provide a neural substrate for learning domain-general mental representations of an animal's environment (O'Keefe and Nadel, 1978), with the entorhinal cortex specifically representing the underlying task structure (Behrens et al., 2018; Whittington et al., 2020).

Connecting these mental representations to behaviour - that is, the actions we carry out - remains an ongoing challenge. Human representations of space typically integrate action-relevant information. This ranges from more implicit information (e.g. cycle paths on a mapping app imply movement by bicycle, and closely-stacked contour lines on a GPS map indicate difficult or impassable terrain) to more explicit information, such as Marshall Islands stick charts, which are designed to show navigators how to adjust their sailing course as a function of changing environmental conditions (Fernandez-Velasco and Spiers, 2024) Nonetheless, it remains to be seen if human *cognitive* maps integrate action information.

In this thesis, I will argue for an inbuilt action representation in representations of relational structure in the entorhinal cortex. In Section 1.1, I will provide a general overview of spatial cell types in the hippocampus and entorhinal cortex. Whereas the hippocampal representations remap easily, those in the entorhinal cortex do not. Some accounts have interpreted this as the maintenance of a consistent representation of task structure across different stimulus sets (Whittington et al., 2020). However, in Section 1.2 I will highlight that the link between entorhinal representations and behaviour in spatial navigation remains unclear. In Section 1.3, I will then discuss how these entorhinal cell types may play comparable roles in visual and abstract spaces, despite lesion evidence questioning their functional role. These various features may be brought together by models of the hippocampal formation which suggest the entorhinal cortex represents a state-action graph (Stachenfeld et al., 2017; Whittington et al., 2022), as discussed in Section 1.4. Nonetheless, the action component of these models is still little explored. Therefore, in Section 1.5 I will review experimental and theoretical work linking entorhinal representations to action and argue in favour of entorhinal cortical responses as a representation of action-relevant space.

On this basis, in Chapter 2 I will present experimental work which tries to resolve the question of whether the human entorhinal cortex represents possible actions in a conceptual space. To test this, we designed an fMRI experiment which teaches participants state-action bindings. The entorhinal cortex represented states in terms of action similarity. Moreover, these representations were tied to eye movement patterns, lending weight to a linking role of eye movements.

Chapter 3 presents a joint fMRI and eyetracking experiment I designed with the aim of elucidating the nature of these action representations via a visual search paradigm. I will present preliminary evidence which finds comparable action representations in a visual space, suggesting that action representations may span multiple domains. Moreover, action representations rotate across task contexts, which may be seen as mirroring rotations previously observed in non-human primate and rodent intracranial work.

Finally, a note on terminology. The terms ‘action’ and ‘affordance’ are used here to refer to domain-general processes which transition (or afford transitions) between states, in line with the modelling and theoretical literature outlined in Section 1.4 (Sutton & Barto, 1998; Dayan, 2012; Behrens et al., 2018). Nonetheless, it should be noted that these terms are generally used to refer specifically to motor action or interactive sensory perception, with a vast accompanying literature (Gibson, 1979). To distinguish these uses, where any mentioned work is specific to the sensorimotor domain I will use ‘motor action’ or ‘perceptual affordance’.

1.1 Spatial representations in the entorhinal cortex

In this section, I will describe how the discovery of spatial cell types in the hippocampal formation has led to the proposal that the entorhinal cortex represents and generalises relational information. Then, I will outline the proposed functional distinction between the entorhinal cortex and hippocampus, whereby the entorhinal cortex may maintain a consistent representation of task structure across different stimulus sets.

For most animals, successful navigation is key to survival (Ekstrom et al., 2018; Yong, 2022). Many species have developed remarkable, highly specialised systems for navigation, such as magnetic field detection in sea turtles and migratory birds (Lohmann, 2010), echolocation in marine or nocturnal species (Moss et al., 2023), or even infrared sensing to find other animals (Campbell et al., 2002). In humans, the visual system allows us to perceive long distances and therefore identify and plan out movements well in advance (Ekstrom et al., 2018). Some navigational feats are so extraordinary that we still don't have a clear understanding of how they are possible, such as salmon and eel returning to breeding sites (Yong, 2022).

The striking diversity of adaptations across species to resolve a single problem ('where am I?') hints at the multidimensional nature of navigation problems. Figuring out position can be achieved via a range of different senses and in self-orientated (egocentric) or globally oriented (allocentric) reference frames (Burgess, 2006). Some tasks require a very low level of positional information or can be achieved by simple stimulus-response patterns, such as directed movement towards a stimulus. Others may require complex, multi-stage planning drawing on memories of previous experiences and generalisation to new sensory stimuli, such as orienting oneself in a new restaurant or migrating across continents. This variation can be seen across species, too: the way a rodent typically navigates, using whiskers for frisking and lateralised vision for orientation, is very different from human long-distance, binocular vision.

In this context, it may seem counter-intuitive to look for shared neural mechanisms across tasks and species. Nonetheless, most navigation problems rely to some extent on understanding where an animal is relative to the surrounding environment, and how this position changes as a function of the agent's action (or external forces). This process, which will be discussed further in Sections 1.4 and 1.5, is often referred to as path integration, sometimes specifically with reference to positional updating in the absence of external cues (Etienne & Jeffery, 2004; McNaughton et al., 2006; Bush et al., 2015).

Identifying where an animal is located appears to be handled, in part, by neural representations of allocentric spatial variables in the hippocampal formation. To unpack a little, this means that pyramidal and stellate

neurons in the hippocampal formation and connected regions release action potentials in response to features of allocentric space.

Among the first spatially tuned neurons to be discovered were place cells, located in the hippocampus, which fire in response to specific locations within an environment during free navigation (O'Keefe & Dostrovsky, 1971; O'Keefe & Nadel, 1978). These were first identified in rodents using electrophysiology, but comparable representations have since been discovered in bats (Ulanovsky & Moss, 2007; Sarel et al., 2017), non-human primates (Courellis et al., 2019), birds (Payne et al., 2021; Agarwal et al., 2023), fish (Yang et al., 2024), humans (Ekstrom et al., 2003) and even sheep (Perentos et al., 2022). In rodents, positional representations seem tied to a specific context: changing environments or goal locations changes how individual place cells map to positions, a process referred to as 'remapping' (Muller & Kubie, 1987). Indeed, place cells have also been found to encode a variety of non-positional variables, such as the direction of movement (Mehta et al., 1997, 2000), location of goals (Ormond & O'Keefe, 2022), upcoming choices (Frank et al., 2000; Wood et al., 2000), social identity (Danjo et al., 2018; Omer et al., 2018) or place-like encoding for pitch (Sakurai, 2002; Aronov et al., 2017). This wide range of encoded variables is consistent with early accounts of place cells and other spatial cell types as the foundation of a 'cognitive map', a domain-general organisation of relational information (Tolman, 1948; O'Keefe & Nadel, 1978; see next section for a discussion on this term).

One early puzzle in spatial navigation was to understand how animals could extract allocentric spatial information from (egocentric) sensorimotor systems. It has been suggested that spatial transformations may be facilitated by head-direction cells, initially discovered in the presubiculum but since found in other regions throughout the medial temporal lobes, including the entorhinal cortex (Taube et al., 1990; Sargolini et al., 2006; Taube, 2007; Giocomo et al., 2014), which encode the heading direction of the animal in relation to an allocentric map. In some cases this is aligned to local or global cues, for example visual cues in rodents (Taube et al., 1990; Bicanski & Burgess, 2016) or magnetic north in shearwaters (Takahashi et al., 2022). This may allow an animal to anchor their own perspective in a global map (Sharp et al., 2001; Bicanski & Burgess, 2018). Head-direction cells can be understood in terms of a circular attractor network, whereby an anatomically-wired ring of neurons encodes head direction using an activity bump which moves smoothly around the ring as a function of idiothetic signals (Redish et al., 1996; Zhang, 1996; Lindsay, 2021; Khona & Fiete, 2022).

Since the discovery of place cells and head-direction cells, many other allocentric cell types have been isolated in the hippocampus and surrounding regions. These include cells which encode the vector distance

to boundaries (O'Keefe & Burgess, 1996; Barry et al., 2006; Bicanski & Burgess, 2020) or objects (Høydal et al., 2019; Andersson et al., 2021), borders (Solstad et al., 2008), goal location (Ormond & O'Keefe, 2022), landmark position (Deshmukh & Knierim, 2013), corners (Sun et al., 2024) or conjunctive combinations of other cell types (Sargolini et al., 2006; Omer et al., 2018). These cells are typically characterised by one or few selective firing fields.

A broad functional distinction has been noted between cells in two of these different regions: whereas in the hippocampus, cells tend to remap freely in response to changes in external stimuli or task condition, in the entorhinal cortex, cells tend to be more stable across environments. For example, both regions contain cells which fire at a given distance and direction to objects. However, landmark-vector cells in the hippocampus often remap as landmarks change (Deshmukh & Knierim, 2013), whereas object-vector cells in the entorhinal cortex tend to maintain their vector firing relationships in the case of moved objects or objects with changed visual and physical features (Andersson et al., 2021). An influential model of episodic memory has argued that spatial or contextual memory information is processed in the medial entorhinal cortex, before being combined with stimulus-specific information drawn from the lateral entorhinal cortex in the hippocampus (Manns and Eichenbaum, 2006). In recent years, this has been extended into the proposal that the entorhinal cortex abstracts and generalises the relational structure of tasks across different sensory stimuli (Whittington et al., 2020, 2022).

Although stimulus-independent encoding of spatial information had already been observed in the entorhinal cortex (Frank et al., 2000), the discovery of grid cells was instrumental in suggesting a role for the entorhinal cortex in generalisation of relational information (Hafting et al., 2005; Fyhn et al., 2007; Bush et al., 2015; Behrens et al., 2018; Bellmund et al., 2018; Bottini & Doeller, 2020; Whittington et al., 2020). These neurons, originally found in layer II of the rodent medial entorhinal cortex, fire for multiple locations in a regularly-spaced hexagonal grid, almost in line with initial implicit predictions of a squared grid based on an attractor network model (Samsonovich & McNaughton, 1997; McNaughton et al., 2006). They are arranged in modules, within which each grid cell shares the same orientation and spacing. The spacing of modules increases along the dorsoventral axis in rodents (Brun et al., 2008). The combination of grid responses across modules allows accurate estimation of the distance and direction between relative positions, thereby enabling path integration (Fuhs & Touretzky, 2006; Stemmler et al., 2015).

As mentioned above, entorhinal cell firing patterns are remarkably stable. When an animal is moved to a new environment, grid cells or object-vector cells may realign (e.g. rotate or translate) in line with new environmental features but maintain the same metric distance information (Fyhn et al., 2007; Høydal et al.,

2019). Therefore, grid cells allow the same distance code to be used across different environments. This invariability can be described with reference to a toroidal manifold, another prediction emerging from continuous attractor network models (Samsonovich & McNaughton, 1997; Fuhs & Touretzky, 2006; Gardner et al., 2022; Guanella et al., 2007; McNaughton et al., 2006; Khona & Fiete, 2022). The cross-cell relationships which generate this torus are maintained across environments, in the dark and even during sleep (Gardner et al., 2022). Therefore, entorhinal cells appear to offer a stable substrate for relational information across different environments or contexts.

In sum, the hippocampal-entorhinal system contains a wide range of cells which represent spatial variables. Whereas hippocampal spatial cells remap easily, medial entorhinal cells tend to realign or rotate rather than fully remap. This has led to suggestions that the entorhinal cortex could provide a neural substrate for learning and abstraction of relational information across different sensory stimuli.

1.2 Spatial cells, action and behaviour

As noted in the introduction, however, it is unclear if and how entorhinal representations are tied to behaviour. While it should be noted that navigational behaviour is complex and is unlikely to show a one-to-one mapping to spatial cell responses (Ekstrom et al., 2020), some indication may be offered by distortions of spatial representations or lesion evidence in rodents. In this section, I will describe how entorhinal cell distortions have led to suggestions that the entorhinal cortex represents action distributions and outline difficulties in connecting spatial representations to behaviour in rodents and humans.

Although entorhinal cells offer a stable coding scheme in comparison to hippocampal cells, it has been observed in rodents that the regular hexagonal firing pattern of grid cells changes as a function of the environment (Barry et al., 2007; Carpenter et al., 2015; Krupic et al., 2016). For example, in a trapezoid arena grid fields are more densely packed in the smaller half (Krupic et al., 2015), an effect which has been linked to spatial memory in humans (Bellmund et al., 2020). In multicompartment arenas with interconnected chambers, grid fields are misaligned when the chambers are separated and then become more aligned over time once they are reconnected (Carpenter et al., 2015). In hairpin mazes, meanwhile, grid cell firing fields split to fire at the end of corridors (Derdikman et al., 2009; Whitlock et al., 2012), a result which has also been reproduced in humans using fMRI and virtual reality (He & Brown, 2019). Distortions also happen in the presence of goals and landmarks, leading to increased firing rates (Butler et al., 2019), which has also been observed for humans in a conceptual space (Viganò et al., 2023), or

realignment towards the goal (Boccara et al., 2019) or visual landmarks (Wen et al., 2024). In humans, it has also been shown that hexadirectional symmetry is disrupted in blind populations, presumably due to different behavioural patterns (Sigismondi et al., 2024).

These distortions appear to arise because entorhinal representations are influenced by features of the environment, although the exact mechanism remains unclear as changes in action distribution, salient locations or local mappings could all explain distortions (Jeffery, 2021; Ginosar et al., 2023). To accommodate the variations in entorhinal responses across different environments, theoretical work has proposed that the hippocampal-entorhinal system plays a role in learning afforded state transitions (Behrens et al., 2018; Jeffery, 2021; Summerfield, 2022). In this paradigm, distortions arise because the agent learns that different positions imply different actions (which can thus be used to guide future behaviour). For example, splits in grid fields in hairpin mazes as a function of direction could indicate that turns should be taken (Derdikman et al., 2009), while increased firing near rewards indicates that the agent should head in that direction (Butler et al., 2019). The entorhinal cortex specifically has been proposed to learn a state-action graph (Whittington et al., 2020), as mentioned in the previous section, or low-dimensional projection thereof (Stachenfeld et al., 2017). Changes in grid cell mappings as a function of environmental change, therefore, would reflect the change in afforded action distribution created by the new context.

The above account may suggest a tight link with entorhinal cell responses and enacted action – i.e. behaviour. Nevertheless, it is unclear to what extent and how entorhinal cell distortions are functionally connected to navigational behaviour, both due to the complexity of navigation problems (Ekstrom et al., 2020) and cases of intact navigation despite medial temporal lobe lesions in humans (McAvan et al., 2022). In a recent experiment, rodents ran in a 1D virtual maze with visual landmarks which predicted reward location. In line with previously observed distortions, grid cells aligned with these landmarks. When landmark position was shifted forward in the maze, both grid cells and behaviour changed, but at different timescales, indicating that there was no one-to-one mapping between grid fields and behaviour (Wen et al., 2024). In other work, however, it has been seen that entorhinal cells predict trajectories that rodents will take in a multi-armed mazes, suggesting that entorhinal representations can indeed predict behaviour (Frank et al., 2000; Gonzalez & Giocomo, 2024; Jones et al., 2024). Moreover, lesions of the rodent medial entorhinal cortex elicit deficits in path integration (regardless of whether navigation is carried out using internal or external cues), suggesting that the entorhinal cortex is functionally involved in the accumulation of movement vectors (Save and Sargolini, 2017).

Therefore, the picture is somewhat mixed: entorhinal spatial representations are altered by changes in an animal's environment, and this can be explained with reference to learning of action distributions. However, establishing a direct link to behaviour has remained difficult. It should be noted, moreover, that the above evidence is largely based on rodent studies and focuses on a single system. While it appears that key features of the hippocampal-entorhinal system are maintained across species (e.g. grid cells; Doeller et al., 2018), there are important cross-species differences (such as more reliance on vision in primates) and both rodent and human behaviour is mediated by a whole range of other systems (Ekstrom et al., 2020).

1.3 Entorhinal representations beyond navigation

In the previous section, I mentioned that entorhinal cell responses can be described with reference to learning of a state-action graph. This perspective may help to explain the involvement of the human entorhinal cortex in both spatial and non-spatial navigation-like tasks, as well as statistical and association learning. In this section, I will discuss evidence for the idea that the entorhinal cortex learns domain-general relational information. As recent work has often interpreted this in light of cognitive maps, I will first discuss cognitive maps and sketch a definition of this term.

A cognitive map, broadly, refers to a mental representation of an animal's environment (Tolman, 1948; O'Keefe & Nadel, 1978; Behrens et al., 2018; Bellmund et al., 2018). Early behavioural evidence for cognitive maps came from the ability of animals to path integrate (Tolman, 1948). Path integration refers to the ability to find new routes in an environment through the accumulation of movement vectors, a process historically described by sailors as 'dead reckoning'. This ability is widespread in non-human species, and is often attributed to humans (Etienne & Jeffery, 2004; Ekstrom et al., 2018). Although cognitive maps have been historically associated with the hippocampus (O'Keefe & Nadel, 1978), in recent years the entorhinal cortex (and in particular the grid network) has been proposed as a neural substrate for path integration (McNaughton et al., 2006), an idea supported by specific deficits in path integration following entorhinal lesions in rodents (Save and Sargolini, 2017) but with diverging evidence in humans – focal entorhinal lesions are vanishingly rare, and hippocampal amnesiacs appear to show intact spatial memory for locations (Rosenbaum et al., 2000) and mapreading (Urgolites et al., 2016), but some deficits in direct pathfinding (McAvan et al., 2022), especially using less-visited routes (Maguire et al., 2006). Indirect evidence comes from a link between impaired path integration and Alzheimer's disease (Segen at al., 2022; Newton et al., 2024), which exhibits entorhinal damage in early stages of the disease (Braak and Braak, 1991).

The term ‘cognitive map’ has been widely used, often with different underlying assumptions. For Tolman the term had a broad use as a mental representation of the environment in any domain which affords flexible behaviour (to the extent that his original work proposed wars were caused by unbalanced cognitive maps; Tolman, 1948). Later accounts in the rodent literature tied cognitive maps to spatial representations in the hippocampus (O’Keefe & Nadel, 1978). This provided a potential explanation for the dual roles of the hippocampal system in navigation and memory (Scoville and Milner, 1957; Buzsáki & Moser, 2013; Eichenbaum & Cohen, 2014), as both mnemonic and spatial information could be organised using the same ‘mapping’ system (O’Keefe & Nadel, 1978). While the *hippocampal* cognitive map is intended as a domain-general representation of information, it should be noted that it is closely tied to space: O’Keefe and Nadel proposed that the representational format was strictly Euclidean and low-dimensional in nature (O’Keefe & Nadel, 1978; for a criticism of this view see Eichenbaum et al., 1999), and work from the Moser lab has largely centred around a better understanding of the spatial positioning system in the hippocampus and entorhinal cortex (McNaughton et al., 2006), occasionally arguing explicitly that other cognition is built upon this system (Buzsáki & Moser, 2013; Bellmund et al., 2018; for a computational take on this see Chandra et al., 2025). Other modern computational work based on reinforcement learning, however, has a distinctly ‘Tolmanesque’ turn, whereby the key goal is to learn and exploit domain-general relational structures, with a particular focus on generalisation (for a review of models see Whittington et al., 2022). Some of these models will be discussed in detail in the next section, but it should be noted that future use of the term ‘cognitive map’ in the thesis will refer to this last definition: an organisation of knowledge into a relational structure that can afford flexible behaviour (Whittington et al., 2022). In all the above cases, however, cognitive maps (hippocampal or otherwise) have been proposed to extend beyond spatial cognition into more abstract domains.

Early studies aiming to show the domain-generality of cognitive maps aimed to teach animals a low-dimensional conceptual space and then show similar ‘space-like’ responses in the hippocampal-entorhinal circuit, in particular place cells and grid cells (Bellmund et al., 2018). For instance, in one task rats were taught to associate specific tones with a reward. When continuous sounds were played, cells in the hippocampus fired for a specific pitch. In the entorhinal cortex, cells fired for more than one pitch, potentially in line with the multiple firing fields of grid cells (Aronov et al., 2017).

A wide range of work has now demonstrated that entorhinal representations code for non-spatial domains. In humans, a landmark neuroimaging study showed that if humans are taught two arbitrary dimensions, in this case the length of necks and legs of bird symbols, transformations in this space elicit a grid-like signal

in fMRI similar to that elicited during virtual navigation (Doeller et al., 2010; Constantinescu et al., 2016). More recent work has extended this finding to a variety of other dimensions, such as social spaces (Park et al., 2021; Liang et al., 2024) discrete semantic spaces (Viganò & Piazza, 2020, 2021), motor action plans (Barnaveli et al., 2025), value spaces (Nitsch et al., 2024), imagined navigation (Sigismondi et al., 2024), sound spaces (Sigismondi, 2024) and even to non-human primates in a value space (Veselic et al., 2025). Moreover, grid-like responses appear to cluster around areas of conceptual space close to goals, in line with grid cells observed during rodent navigation (Viganò et al., 2023). Incredibly, recent intracranial work has even demonstrated that human entorhinal neurons fire in a hexagonal pattern within a two-dimensional space defined by valence and arousal in an image recall task, even though these dimensions were never explicitly probed (Qasim et al., 2023).

Human neuroimaging work has also shown that spatial distances are represented in the entorhinal cortex during virtual reality navigation (Howard et al., 2014). Distance codes have also been observed in conceptual spaces within both the hippocampus and entorhinal cortex, in some cases combining or selecting multiple learned dimensions (Theves et al., 2019; Park et al., 2020; Theves et al., 2020; Viganò and Piazza, 2020, Sweigart et al., 2023). This goes in line with a role for the entorhinal cortex in learning relational information across domains.

A related line of work has extended entorhinal representations to visual spaces. In primates, vision is a much more dominant sense than in rodents (Nau, Julian, et al., 2018). Vision is an especially interesting domain here, as it bridges both spatial and non-spatial exploration: changes of state in conceptual spaces can elicit accompanying eye movements (Hartmann, 2022; Viganò et al., 2024). Moreover, eye movements are often taken as a read-out of both external and internal attention (Groner & Groner, 1989; Hoffman & Subramaniam, 1995; Deubel & Schneider, 1996), which has a close link to memory (Chun & Turk-Browne, 2007; Muzzio et al., 2009; Meister & Buffalo, 2016; Nau, Julian, et al., 2018; Bottini & Doeller, 2020; van Ede & Nobre, 2023). In primates, intracranial work has demonstrated both grid cells in visual spaces which are modulated by familiarity (Killian et al., 2012; Killian & Buffalo, 2018) and specific saccade-direction cells in the primate entorhinal cortex, which appear to play a similar role to head-direction cells or vector cells (Killian et al., 2015). This has been backed up by both intracranial electrophysiological and non-invasive work in humans, which has shown that eye movements elicit a grid-like signal (Julian et al., 2018; Nau, Navarro Schröder, et al., 2018), in some cases also modulated by familiarity (Graichen et al., 2024). Whether these visual responses are dissociable from attention remains an open question. In an MEG study, Giari and colleagues showed using a frequency-tagging paradigm that movements of covert attention elicit increased signal at 60 degree angles, similar to the hexadirectional signal previously observed in fMRI and

iEEG (Giari et al., 2023). This is in line with previous work in non-human primates suggesting that grid response can be elicited by covert attention (Wilming et al., 2018).

Beyond these conceptual and visual spaces, fMRI studies have indicated that the entorhinal cortex also has a role in sequence memory (Bellmund et al., 2019, 2020) and association or statistical learning both in humans (Klingberg et al., 1994; Hindy and Turk-Browne, 2015; Baram et al., 2021) and non-human primates (Buckmaster et al., 2004; Garcia and Buffalo, 2020; Ku et al., 2021). For example, Shpektor and colleagues (2024) taught participants to associate images with musical notes in an auditory sequence with a hierarchical structure. When the representations of these images were subsequently tested using fMRI, it was observed that neural pattern distances in the entorhinal cortex scaled with the hierarchical distance of objects in the auditory sequence (Shpektor et al., 2024). These findings are consistent with a role of the entorhinal cortex which extends beyond a mapping of spatial location to the construction of more complex relational structures, which has led to suggestions of a general learning mechanism for cognitive maps in the entorhinal cortex (Whittington et al., 2020). If the entorhinal cortex is recruited to learn complex associations rather than specialising in a particular domain, this may also help to explain potentially conflicting results such as preserved spatial (McAvan et al., 2022) or semantic memory (Tulving et al. 1991; Knowlton and Squire, 1993) following medial temporal lesions in rodents and humans.

This may also be applicable to the motor domain, where amnesiacs with entorhinal damage appear able to learn some motor tasks (Shadmehr et al., 1998; Squire, 2009) but struggle to grasp more complex action-outcome associations (McDougle et al., 2022). This will be discussed more fully in the thesis conclusion in relation to our findings, but would be consistent with a role for entorhinal cortex in learning more complex graphs based on action associations, as will be discussed in the following section.

Nonetheless, it should be noted that there is still little evidence linking associative learning in the entorhinal cortex with the conceptual and visual “navigation-like” representations described above (Bellmund et al., 2018). Human fMRI studies have shown that step distances in complex graphs are be encoded in the entorhinal cortex (Garvert et al., 2017), as well as hierarchical sequences (Shpektor et al., 2024), which suggests that the entorhinal cortex does extend associative learning to multiple steps. However, it is unclear how these representations relate to findings of grid-like signal in conceptual spaces: it is possible that they rely on separate functions of the entorhinal cortex as either a spatial ‘scaffold’ (Chandra et al., 2025) or associative memory learner (Buckmaster et al., 2004) in different task contexts, rather than drawing on the same general learning system (Behrens et al., 2018; Whittington et al., 2020).

In sum, these findings suggest that neuronal populations in the entorhinal cortex may code for relational information which extends beyond real space to conceptual and visual spaces. This role in general-purpose learning is in line with theories predicting that the entorhinal cortex learns relational structures (Whittington et al., 2020) and activation during sequence and associative learning tasks, although the link between the two has yet to be established. Under some accounts, this may be seen as learning a cognitive map in line with a proposed role in path integration (Behrens et al., 2018). Nevertheless, it should be noted that the evidence presented above is largely correlational, and the wide range of functions and possible compensatory mechanisms would make this theory difficult to falsify. While Behrens and colleagues (2018) propose graph learning as a unifying paradigm, there nonetheless remains an open debate over the format of entorhinal representations (Peer et al, 2021) and it is unclear if positional representations are part of the same entorhinal network as associative learning (Garcia and Buffalo, 2020). Moreover, while previous work has identified representations related to motor action in the human entorhinal cortex (Barnaveli et al., 2025), and oculomotor action in the primate entorhinal cortex (Killian et al., 2015), which will be discussed further in the main conclusion, it remains unclear if the entorhinal cortex indeed learns state-action associations across domains as predicted by reinforcement-learning inspired models. This will be discussed further in the thesis conclusion in light of our findings and the models presented below.

1.4 Models of cognitive maps

Thus far, I described a proposed functional distinction between the entorhinal cortex and hippocampus, whereby entorhinal cortex responses remain ‘stable’ across different stimulus sets (Section 1.1). Nonetheless, entorhinal representations during spatial navigation do appear sensitive to features of the environment, possibly driven by action possibilities (Section 1.2). Entorhinal populations may also code for association learning and other non-spatial domains, possibly as part of a cognitive map (Section 1.3). These ideas have been formalised in computational models, which provide predictions about the type of representation we may find in the entorhinal cortex and aim to explain the entorhinal cortical role in both associative learning and spatial navigation. I will now discuss these models, and highlight how they formalise a link between entorhinal relational structures and behaviour via learning of an action component.

The models of the entorhinal cortex that will be discussed can be divided into two broad categories (Dong & Fiete, 2024). In the first case, models such as the Tolman-Eichenbaum Machine (Whittington et al., 2020) or successor representation models (Stachenfeld et al., 2017) use minimal biological constraints, instead aiming to replicate functional properties of cognitive maps using reinforcement learning. In these cases,

cognitive maps can be state-action graphs able to accommodate any (learnable) information, and only in special cases replicate spatial features of the entorhinal cortex such as the toroidal grid cell manifold or multiple modules (Schaeffer et al., 2022; Dorrell et al., 2022; Dong & Fiete, 2024). On the other hand, continuous attractor network models such as Vector-HASH (Chandra et al., 2025) start from strong biological or functional constraints based on spatial navigation and may then try to understand how these features may be applied to non-spatial domains. This leads to cognitive maps based on continuous, two-dimensional spaces which can map onto a spatial scaffold.

Among other cell types, the models often aim to replicate grid cells, which may not be the most stringent constraint: a hexagonal grid can be explained as the best way to stack circles together (Mathis et al., 2015), which captures, in a loose sense, the logic of continuous attractor network models. Grid fields can be recreated via principal component analysis of place cell firing fields (Dordek et al., 2016), which may themselves be an epiphenomenal feature of any RNN (Luo et al., 2024). Indeed, grid fields can be found in random Gaussian noise after low pass filtering when using a low threshold for grid cell identification (Sorscher et al., 2023). Nonetheless, the clear link to neuroscientific results means that these models may provide predictions at multiple levels of analysis (e.g. biological: grid fields; algorithm: attractor network; computation: path integration).

Successor representation models: learning action for next state prediction

A basic assumption of models based on learning state-action graphs is that the hippocampal formation functions in a similar way to model-based reinforcement learning (RL) models (Sutton & Barto, 1998). In RL, an agent learns how actions can change between states (typically associated with a reward or penalty), such that these transitions can then be used to calculate the expected value of future trajectories. Actions can correspond to any input which changes state: this could mean a movement in space, or a non-spatial action such as increasing in sociability (Dayan, 2012).

In the case of successor representation models (Stachenfeld et al., 2017; Gershman, 2018), a hybrid between model-based and model-free learning, the agent learns state-state relationships through experienced action-based state transitions. Under the assumption that these guide behaviour, this is used to create a ‘successor matrix‘, which predicts the expected future occupancy of each state given the current position and a temporal discount factor. It has been argued that this predicts human decision-making (Momennejad et al., 2017). In a spatial framework, this matrix would assign high value to spatially-adjacent states, or nearby

spatial positions associated with a reward. When projected onto a Euclidean space, this can be used to generate place fields which skew in the direction of movement and remap in the presence of changing environments or reward, just as in real neural data (Stachenfeld et al., 2017). Notably, however, the RL framework means that this approach can be used for any environment which respects the basic constraints of state-action mappings.

To generate entorhinal cell populations, the eigenvectors of the successor matrix are calculated. When mapped onto the original space, some of these eigenvectors resemble grid fields and other entorhinal cell types such as border cells, and capture distortions of grid fields in uneven environments (Stachenfeld et al., 2017). Proponents of this model argue that this means entorhinal cells can be used for error correction or even generalisation (Yu et al., 2021), as they provide a lower-dimensional map from which the original place fields (and thus the environment) can be reconstructed. This represents an elegant explanation for grid cell firing patterns: the grid fields emerge from the even distribution and topology of possible actions and states in spatial navigation. Moreover, this secondary role for grid cells reflects the minimal disruption of grid cell lesions on place cell functioning (Hales et al., 2014). This means entorhinal responses are susceptible to changes in action distribution or environment. For example, in a square environment with a square graph many ‘cells’ appear to have four-fold rather than six-fold symmetry (Stachenfeld et al., 2017). At a population level, therefore, we may expect a neural signal which reflects the task constraints in terms of predicted state-state transitions, which vary as a function of action.

The Tolman-Eichenbaum Machine: learning a state-action graph for latent inference

A related approach to modelling cognitive maps is to build a modular structure into the model which mirrors the biological role of the hippocampus and entorhinal cortex. For example, the Tolman-Eichenbaum Machine (Whittington et al., 2020) draws on the dual hippocampal functions of relational memory and navigation to propose a combined model which stores state information in a ‘hippocampal’ network using Hebbian learning, and predicts new states using a ‘medial entorhinal’ RNN trained on a path integration-like task in a state-action graph, as well as a ‘lateral entorhinal’ associative network which stores state information. This allows the network to perform generalisation by matching stimuli to the relevant state in a previously-learned latent graph, and thereby understanding via transfer how experimenter-provided actions move within the graph.

Nodes in the network generate responses which resemble rate maps from rodent electrophysiology, including place cells, band cells, grid cells and object vector cells. As in the entorhinal cortex, grid cells respond at multiple spatial scales and retain their relationships across different stimulus sets. The hexagonal grid responses are somewhat dependent on parameter choice, in particular regularisation (Schaeffer et al., 2022; Sorscher et al., 2023). An interesting prediction emerges from object vector and border cell representations: these only appear when behaviour is biased toward objects or boundaries, suggesting that they may be tied to behaviour. Indeed, subsequent work has proposed object or goal vector cells as a potential neural substrate for implied (or afforded) action, which will be discussed later (Whittington et al., 2020, 2022). More generally, as with the successor representation model, action is closely integrated into the system as the model learns the action possibilities and outcomes for each state.

In sum, TEM demonstrates that a modular network aiming to solve generalisation problems using a latent state-action graph and memory module replicates key features of entorhinal cell types and generates new predictions for when they should emerge. That said, it is important to note that while graph-based networks such as TEM and SR model the flexibility of cognitive maps and a wide variety of entorhinal cells, they only replicate the invariability of grid cells (i.e. attractor network, fixed correlation structure and resulting toroidal manifold) in certain conditions (Dong & Fiete, 2024). Moreover, some elements which may be well-defined in some reinforcement learning contexts or spatial navigation may remain less clear in abstract contexts, such as the action input, which lacks a well-established neural substrate. Therefore, an alternative approach to building a cognitive map may take the opposite route: to start from a model which is able to carry out path integration in a Euclidean space and extend that to non-spatial domains.

Continuous attractor network models: action as velocity input to the grid network

As noted in Section 1.1, grid cell responses can be described with reference to a toroidal manifold (Gardner et al., 2022). This was predicted by continuous attractor network (CAN) models (Samsonovich & McNaughton, 1997; McNaughton et al., 2006; Burak & Fiete, 2006; Fuhs & Touretzky, 2006; Guanella et al., 2007; Burak & Fiete, 2009; Khona & Fiete, 2022). In these paradigms, grid cell responses are generated due to local inhibition and global excitation between grid cells, where local refers to closely-wired neurons. This generates an activity bump with a characteristic ‘Mexican hat’ shape whereby neighbouring cells are less likely to fire. The multiple firing fields of grid cells are generated because of periodic boundary conditions, which means that a movement of one phase length (a lattice vector) brings the activity bump back to the same neuron. The resulting rhomboid can be folded into a toroidal manifold, whereby each

individual grid cell describes a single point on the 2D manifold. A grid module with the same spacing and orientation, therefore, can be described as a set of points on a toroidal manifold (McNaughton et al., 2006).

Models based around continuous attractor networks, therefore, start by metaphorically hard-wiring the grid module architecture which leads to the formation of a torus, before using neural networks to learn how torus positions are associated with real space or memory. There is evidence from electrophysiology that suggests the torus may indeed be hard-wired, rather than learned: even if rodents are not exposed to flat spaces during development, grid cells nonetheless emerge during adulthood (Ulsaker-Janke et al., 2023). The base assumption of CAN models, expressed via this constraint, is that relationships in cognitive maps can be understood as positions on low-dimensional tori. This does not necessarily limit cognitive maps to few dimensions, as higher dimensions can be composed of multiple tori (Iyer et al., 2024), but does imply a continuous map. One advantage of this approach relative to graph models is that new actions can be immediately carried out without learning: any movement in torus-space is possible from any position.

In spatial navigation, activity bumps are moved around by the input of velocity signals, which are derived from self-motion signals such as idiothetic information. In an influential model of spatial navigation, these are derived from parietal regions in an egocentric format, before being converted into an allocentric framework and projected to the medial temporal lobes (Bicanski & Burgess, 2018). Self-motion-related signals have been identified throughout this processing stream, from parietal area 7A (Siegel & Read, 1997; Avila et al., 2019; Alexander et al., 2022) to the entorhinal cortex itself (Mallory et al., 2021). One challenge for continuous attractor models aiming to explain navigation in non-spatial dimensions such as Vector-HASH (Chandra et al., 2025) is accounting for how we move between positions. If the positions themselves are tied to known memories or concepts, velocity input can in theory be calculated via retrieval and subsequent trajectory calculation (with reference to the path integration function of grid cells). But in order to retrieve new memories, we need a mechanism to store and implement these velocity signals. In Vector-HASH, these velocity inputs are learned and stored by a feedforward network in the hippocampus, but this remains an open experimental question. Therefore, whereas the previous models suggest that the entorhinal cortex should be susceptible to changes in action distribution, CAN models are less clear on how action is processed in a conceptual space.

In sum, models of hippocampal-entorhinal function based on graph learning or continuous spaces pose different solutions to how a cognitive map is learned, but also different hurdles. Models strongly grounded in spatial navigation provide stronger constraints on the type of information that may be encoded, and therefore clear predictions arising from spatial navigation. Nonetheless, it is not entirely clear how spatial

frameworks map to abstract knowledge, for example in the case of velocity input, unexplored areas of memory or high-dimensional information. Models based on graph learning, on the other hand, offer important bonuses in terms of flexibility and domain-generality. Nonetheless, they require extensive learning of new environments and may only replicate biological data in special cases. However, all these models can replicate the relative stability and domain-generality of entorhinal representations. In terms of our main question, graph learning-based models predict an action representation in the entorhinal cortex, while this remains an open question for continuous attractor network models.

1.5 A missing link to action

In the previous sections, I have outlined how experimental data suggests a role for the entorhinal cortex in mapping out relational information across domains. In spatial navigation, entorhinal representations are sensitive to environmental changes, which is captured by some modelling work as the learning of state-action dependencies. If this is indeed the case, the entorhinal cortex should learn and represent actions within relational structures. Although there is some evidence that the entorhinal cortex may be involved in learning of associations and motor action contingencies, it is unknown if this is reflective of a general role in action learning beyond the motor domain. Based on this, I will aim to test if the human entorhinal cortex represents possible actions in conceptual and visual spaces.

The models presented above have presented two modes of representation for action: as a separate input provided by another other system (as in self-motion signals during spatial navigation), or alternatively as a feature of learned relational information (for example, distortion of spatial cell types as a result of environmental change). In general, it is difficult to distinguish between the two types of action representation. For example, apparent distortions in spatial representations could be driven by either a change in entorhinal representation or modification of velocity signal - which might indeed be a more flexible candidate (McNaughton et al., 2006) and may also be influenced by environmental geometry (Munn et al., 2020) or Alzheimer's disease (Ying et al., 2023). To take an example, in an experiment from Neupane and colleagues, it was found that entorhinal cells fired at stimulus-aligned peaks during a mental navigation task in which monkeys had to move a joystick to move between imagined animal images (Neupane et al., 2024). This could be explained either by an alignment of entorhinal cells to landmarks, or a modification of velocity signal relative to hand motion such that the cell periodicity matched the stimulus position. A second difficulty in testing the origin of action representations experimentally is due to imagination. Rodents can remotely activate place fields (Lai et al., 2023) and imagined navigation elicits

periodic responses in the primate entorhinal cortex (Neupane et al., 2024). If imagined action can elicit similar neural signatures to enacted action, then it may be difficult to distinguish an action-affording environment (e.g. passive interaction with a state) from action input (e.g. changing state) using macro-level techniques such as fMRI. A final obstacle is that the models discussed above do not offer clear predictions: for example, TEM and Vector-HASH can both learn state-action combinations and take actions as input. Due to these considerations, we do not aim to distinguish between action input and an action-affording representation. Indeed, as prior literature has reported both self-motion signals in the entorhinal cortex (Mallory et al., 2021) and environmental distortions (Ginosar et al., 2021), it is possible that any observed action representation is a mix of both signals.

As outlined previously, it has been proposed that the entorhinal cortex generalises task structure across different stimulus sets. This may extend to actions, which may elicit the same neural responses both within and across relational structures. The authors of TEM have proposed goal-vector cells, a behaviourally-modulated version of object-vector cells, as an underlying neural substrate which can 'compose' actions within and across maps (Whittington et al., 2020, 2022). This is a neat solution, as these cell types can effectively encode vectors in the same space as the learned graph, and has other benefits in terms of generalisation and composition when paired with replay (Bakermans et al., 2025). Thus far, neural evidence for this proposal is lacking. While we are unable to directly test the neural populations involved due to the spatiotemporal limitations of human neuroimaging, it may be possible to tap into the broader question: do action representations remain consistent across different task contexts? I will design an experiment to test this in Chapter 3, where the experimental background and hypotheses are laid out more fully.

Thus far I have largely focussed on entorhinal representations of action, but it is likely that other systems are responsible for coordinating and integrating action information into the medial temporal lobes during non-spatial tasks, just as in spatial navigation (Bicanski & Burgess, 2018). For instance, it has been shown that movements of attention (both covert and overt) both generate entorhinal responses (Ringo et al., 1994; Julian et al., 2018; Killian & Buffalo, 2018; Wilming et al., 2018; Giari et al., 2023) and are invoked during exploration of conceptual spaces. The most stark example of this comes from numbers: when participants are asked to retrieve random numbers while looking at a blank screen, their eyes shift leftward for smaller numbers and rightward for larger numbers (Loetscher et al., 2008, 2010). This has since been observed in other conceptual domains (Viganò et al., 2024), and it has been argued that oculomotor mechanisms, just like hippocampal-entorhinal mechanisms, could play an important role in the 'navigation' of our memory (Meister & Buffalo, 2016). In light of this, we might expect regions which handle oculomotor and attentional signals, such as the posterior parietal cortex, to play a role in integrating action information into

cognitive maps. Suggested candidates for integration of action, too, such as vector cells and hippocampal cells (Chandra et al., 2025), which indicate the location of behaviourally-relevant targets such as goals (Crivelli-Decker et al., 2023; Ormond & O'Keefe, 2022), may also be responsive to oculomotor or attentional input (see Chapter 2 for a discussion of saccade direction cells in this light).

More generally, it should be observed that relatively little is known about the relation between entorhinal representations of relational structures and behaviour. Although I have presented evidence that entorhinal spatial representations appear to be responsive to and modulated by action-related features such as environmental shape or goal location, the general consensus is that the entorhinal system, unlike the hippocampus, is remarkably resilient to external changes (Derdikman et al., 2009) and does not immediately map to behaviour at rapid timescales (e.g. Wen et al., 2025). Integration of action into entorhinal representations may explain experimental phenomena and facilitate exploitation of relational information (e.g. Bakermans et al., 2025), but this does not imply a one-to-one link between entorhinal representations and behaviour – for example, it may be advantageous to prioritise exploration over exploitation (Sutton & Barto, 1998; Gershman & Daw, 2017). Moreover, there may be multiple explanations for experimental phenomena which could combine to drive behaviour: a map of action-relevant space might be compatible with other explanations for grid cell distortions such as local metrics or a mapping of subjective space (Ginosar et al., 2023). That said, understanding how the entorhinal cortex represents action information may bring us closer to understanding how the human brain goes from learning relational structures to enacting the aforementioned actions.

In light of this, the current thesis will attempt to understand if the entorhinal cortex represents action, in the eventual hope of linking structured representations to behaviour. To do so, we designed two fMRI and eyetracking studies aiming to isolate action representations in the entorhinal cortex. The two key questions we ask are the following:

Do entorhinal relational maps contain action representations across domains? In Chapter 2, I designed and ran an fMRI experiment in which participants were taught to associate numerically-labelled states with mathematical operations, as a proxy for action in an abstract space. I find that the right entorhinal cortex represents states in terms of action similarity. This is backed up by preliminary evidence presented in Chapter 3, in which a similar paradigm is used whereby participants learn to associate animal images with eye movement possibilities.

Are entorhinal representations of action context-specific? Previous theoretical work has suggested that action representations may constitute a separable neural substrate. In Chapter 3, I aim to test this by comparing action representations across two graph contexts. We found that neural patterns correlate best with an action matrix rotated across contexts.

Representation of conceptual affordances in eye movements and the entorhinal cortex

2.1 Abstract

The hippocampal-entorhinal system represents relations between states in spatial and non-spatial domains. Critical to understanding how these memory representations are used for cognition is to determine whether the actions underlying state transitions are incorporated in entorhinal representations of relational structure. Participants learned to transition between states using different mathematical operations (actions), which they were able to generalise to previously-unseen states. We found that the entorhinal cortex represented the afforded operations across the states. This affordance representation was not explained by other properties of the task space such as link distance between the states, action magnitude or reaction times. Furthermore, gaze behaviour reflected the direction of afforded operations in the horizontal axis, and the strength of this lateralisation predicted performance and entorhinal activation, suggesting a link between gaze behaviour and neurocognitive mechanisms for navigating conceptual spaces. In sum, this study provides first evidence for the integration of action information into ocular and entorhinal representations of conceptual spaces, suggesting that these may not just map out experiences, but provide information about how to explore knowledge.

Note: An earlier version of this chapter is currently available as a preprint on the bioRxiv preprint server (Eperon et al., 2025).

2.2 Introduction

Human representations of space are closely connected to behaviour. Whereas early maps tended to highlight topography useful to hunter-gatherers, modern mapping apps accentuate the differences between major roads and cycle paths. In light of this, how does the brain support the interplay between spatial representations and action?

The mammalian hippocampal formation is known to contain a variety of spatially-tuned cell types which respond selectively to locations in space (e.g. O'Keefe & Dostrovsky, 1971; Hafting et al., 2005; Høydal et al., 2019) and are thought to offer a neural substrate for a cognitive map (Tolman, 1948; O'Keefe & Nadel, 1978). Recent evidence has supported the idea that the hippocampal formation encodes relations in support of memory, across spatial and non-spatial domains (Manns & Eichenbaum, 2006). For instance, studies have observed that during transitions in two-dimensional feature spaces the entorhinal signal shows a hexa-directional modulation by the trajectory angle (Constantinescu et al., 2016; Viganò & Piazza, 2020, 2021; Park et al., 2021; Nitsch et al., 2023; Qasim et al., 2023; Schäfer et al., 2024; Viganò et al., 2024), mirroring the grid pattern observed during spatial navigation (Doeller et al., 2010). Relatedly, the representations in the hippocampal formation reflect the distances between positions in abstract spaces, such as link distance on a graph (Schapiro et al., 2012; Garvert et al., 2017, 2023; Pudhiyidath et al., 2021), distances in conceptual and social spaces (Theves et al., 2019, 2020; Park et al., 2021; Schafer et al., 2022), or location along a sound frequency spectrum (Aronov et al., 2017).

To exploit such relational knowledge representations, however, we need to know how to transition between different states. For instance, to win a chess game, we need to know which actions are afforded at any given state of the game in order to move the pieces from the initial setup to checkmate. In physical navigation, it is thought that self-motion signals are converted into a velocity input which determines movement around the grid attractor network (Khona and Fiete, 2022). In other words, perceptual, vestibular and motor signals provide the direction and speed information necessary to update the agent's position in the entorhinal network. Although there is no clear non-spatial analogue for this process, domain-general models of the hippocampal-entorhinal system often use reinforcement learning terminology to describe how 'actions' are used to change between 'states' (Sutton & Barto, 1998). Accordingly, different models integrate action input as the cue to move between different states of learned graphs (Stachenfeld et al., 2017; Whittington et al., 2020; George et al., 2021) or a pre-structured Euclidean space (Chandra et al., 2025). However, whether and how action information is actually incorporated in relational memory representations remains to be established empirically (Aronov et al., 2017; Sakurai, 2002).

Here we focus on the entorhinal cortex, which has been proposed as a neural substrate for the abstraction and generalisation of task structure, including action information (Behrens et al., 2018). Previous work has shown that entorhinal cortex representations generalise task structure across different stimulus sets (Baram et al., 2022; Mark et al., 2024), and modelling work has suggested that the entorhinal cortex learns the possible actions within these structures (Whittington et al., 2020), potentially as part of a larger system for compositional planning (Bakermans et al., 2024). Nonetheless, experimental evidence for this proposal is lacking. Therefore, we investigate if and how the entorhinal cortex represents action information within learned relational structures.

In the context of spatial navigation, although the exact nature of action representation is unclear (Ginosar et al., 2023), there is increasing evidence that entorhinal representations are sensitive to changes in idiothetic signals such as speed (Kropff et al., 2015, 2021) and gaze and head position (Mallory et al., 2021), as well as behaviourally-relevant environmental changes. Moreover, entorhinal cells are sensitive to changes in landmarks or goal position (Boccara et al., 2019; Butler et al., 2019; Gonzalez & Giocomo, 2022; Wen et al., 2024). This has also been observed in human neuroimaging studies, which have shown distortions of grid-like signatures in fMRI based on goal locations (Viganò et al., 2023), arena layout (He & Brown, 2019) or prior levels of visual experience (Sigismondi et al., 2024). These distortions might be interpreted as a representation of 'action-relevant space' (Ginosar et al., 2023). Moreover, incorporating a condition of 'actionability' (that is, representations must enable actions to be carried out) appears to improve on existing models of grid cells by replicating key features such as multiple modules (Dorrell et al., 2022). Nonetheless, it remains an open question if actions (or possible transitions) themselves are represented in the entorhinal cortex independently of specific stimulus features or environmental layout.

Potential transitions in cognitive spaces have also been shown to be reflected in eye movements, possibly indicating shifts of attention (Loetscher et al., 2008, 2010; Giari et al., 2023; Viganò et al., 2024). Eye movements also elicit similar representations in visual space to navigable space in the entorhinal cortex (Killian et al., 2012, 2015; Nau et al., 2018; Staudigl et al., 2018; Giari et al., 2023; Graichen et al., 2024), and hippocampal place cells in chickadees code for the same location whether visited physically or visually (Payne and Aronov, 2025). Accordingly, we also used eyetracking to evaluate directional eye movements in our task and their relation to neural representations.

In sum, we investigated the neural representation of possible state transitions (affordances) using fMRI and eye-tracking while participants evaluated numerical operations (actions) afforded by given numbers in a

repeating graph structure (states). We found conceptual affordances to be represented in eye movements and entorhinal cortex activation patterns.

2.3 Materials and methods

In the following section, I will describe the experimental design and analyses used to test our primary hypothesis: that the entorhinal cortex represents possible action information in a conceptual space. We aimed to address this by comparing possible action distributions between states in a conceptual space. Firstly, I will briefly touch on the experimental logic which led to the choice of a number-based design.

One key challenge was designing an experiment which could test action independently from other variables which might describe a relational structure. Prior studies of entorhinal representations largely aimed to teach participants an entirely-novel set of stimuli or structure, such as birds with varying leg and neck length (e.g. Constantinescu et al., 2016), or a graph network based on image associations (e.g. Garvert et al., 2017). In our case, however, this posed a problem: to compare affordances across different states, we needed to know if two actions were indeed considered similar or not. This requires an underlying structure which can be used to identify similar or different actions (for example, two movements eastwards are only similar in virtue of a shared reference frame). The problem arose, therefore, of how to teach both an underlying structure and a changing action distribution. In more concrete terms, how could we teach that A->B is the same as C->D without relying on a more concrete domain?

The solution arrived at was to project learned actions onto a known conceptual space with a well-established structure. After some piloting, we decided to use a number line. Participants would learn a trajectory overlaid on the number line (e.g. 2->4->3->5), arranged in repeating graph modules every five numbers. Actions are therefore operationalised as mathematical operations: moving from 2->4 requires the same action as 3->5. This enabled us to carefully control the relation between states and actions, such that action similarities could not be explained by position, numerical distance, link distance, or other potential confounds. Moreover, the repeating structure allowed us to test participants on previously-unseen stimuli, encouraging active engagement with the task structure. Finally, a wide range of work has shown that numerical cognition elicits directional eye movements. Therefore, we could both test engagement with the numerical features of the task using eyetracking (e.g. eye movements should skew further left for ‘negative’ actions than for ‘positive’ actions) and investigate how eye movements interact with entorhinal representations of entorhinal structure.

Having described the logic of the design, I will now describe how we taught action affordances to participants, recorded eyetracking and fMRI data, and subsequently analysed the data to test for a representation of affordances in the entorhinal cortex.

2.3.1 Experimental design and task structure

Experimental design

The experiment took place over two consecutive days. On the first day, participants were informed that they would play the role of an archaeologist who had discovered a new number system in local sandstone caves whereby numbers could be transformed into other numbers. This was, of course, fictitious and designed to ensure they understood that they should try to learn number transitions from scratch rather than apply any pre-existing intuitions. Participants were instructed that their task was to first understand how numbers could 'transform' into other numbers, and then use this knowledge to identify correct number transformations. These transformations would define the ‘actions’ used in future analyses: if the number 25 could transform into the number 27, for example, we can intuit that 25 affords the action +2.

Each day consisted of both training and a recall task. The training was designed to teach participants how numbers were connected to other numbers by interleaved exploration and test blocks (see below). The recall task was used to measure neural and ocular responses to each number, which would then be tested based on the similarity between affordances.

On the first day, participants carried out the training for approximately an hour and a half (40 blocks), followed by half an hour of a recall task (4 runs) while eye movements were monitored using an eyetracker (details below). On the second day, they repeated the training for only 20 minutes (10 blocks), followed by an hour of the recall task (8 runs) while brain activity was monitored using fMRI. At the end of the experiment, participants were asked to fill out a questionnaire, which primarily aimed to test strategies used while carrying out the task, and gather relevant demographic information.

Stimuli

As mentioned above, our design aimed to distinguish between previously-tested features of a state space (e.g. link distance) and the specific transitions (numerical operations, referred to interchangeably as 'actions') that could be carried out within this structure.

Figure 1. *Numerical stimuli were arranged into repeating graph modules (left; the full graph consisted of 144 numbers), which can be described as repeating states (right). This allowed us to manipulate the actions from each number independently from position. Moreover, the repeating modules allowed us to probe the nature of action representation independently of individual number identity.*

To dissociate state identity from specific stimulus identity, we created a repeating structure (module) based on relations between four states (Figure 1). This module repeated every five numbers. Each state (number) was associated with two possible actions (arithmetic calculations:+2,-2,+1,-1) and participants could navigate the number line by choosing a transformation which corresponded to one of them. For instance, if number 23 afforded actions -1/-2, it would be possible to select 21 but not 25 as the next state.

Using a number line as the base structure allowed us to manipulate the transitions independently from the specific states. For example, transitioning between any two numbers could involve a big numerical change but still constitute one link in the graph structure. As previous studies reported a linear encoding of numerical knowledge, we assumed that the relationship between actions would hold independently of the numbers seen (e.g. '3' to '5' would be similar to '55' to '57'). Likewise, while a movement from ‘3’ to ‘2’ may be a single step, it constitutes a different transition (-1) from that of the next step ‘2’ to ‘4’ (+2).

To further control for any pre-learned mathematical associations or visual confounds, we balanced number-state assignments across participants by shifting the position of graph modules along the number line. Whereas one participant may learn the trajectory 25 - 27 - 26 - 28, another might learn 36 - 38 - 37 - 39.

In total, participants were trained and tested on 4 and 8 modules (16 and 32 numbers) respectively prior to entering the fMRI scanner, while each run consisted of an extra 3 previously unseen modules (12 numbers). In total, the available structure seen throughout the entire experiment consisted of 36 modules per participant, corresponding to 144 numbers. All numbers were greater than 20, to avoid interactions with pre-existing associations related to single-digit numbers.

Training

Training (Figure 2, below) was designed to teach participants the transformations between numbers using a 'free exploration' paradigm. Participants were presented with one number on the screen, above two options which corresponded to the two directly connected numbers in the graph structure introduced above. Selecting a number (using a button press) from the two choices led to the next display, which showed the selected number with its two direct connections (see Figure 2). This phase can be seen as an exploration of the number line using the numerical operations available in the structure.

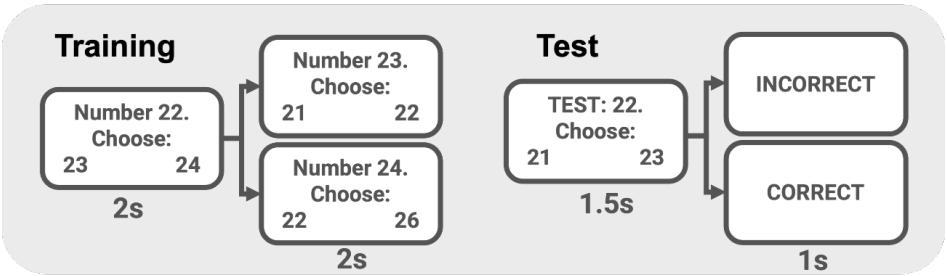

Figure 2. *Participants learned the graph structure via a free exploration paradigm, in which they were first given a free choice of two possible successor numbers according to the graph structure (left). They were then probed on their ability to select the correct successor for each number in a binary choice task (right).*

After 16 exploration trials, participants were presented with 18 test trials on the same numbers, in which only one of the two options corresponded to a correct successor according to the graph structure. In these trials, participants were presented with a similar screen, consisting of a central number above two possible successors. In this case, however, only one successor was correct. Participants were asked to select the correct successor and were given feedback on the screen after each trial.

After 20 and 40 of these train and test blocks, as well as at the end of the training on the second day, 24 test trials using novel numbers (drawn from the beginning and end of the participant-specific number line) were presented without feedback. This allowed a test of how well participants applied the learned transitions to new numbers, which we refer to as 'weak generalisation' due to visual clues resulting from the length-five graph modules (i.e. if number 29 afforded 27 and 31, number 79 would afford 77 and 81).

Test trials were evenly selected from different positions within stimulus modules. Most trials were selected from within a range of two from the presented number in order to avoid use of a distance-based heuristic. Nonetheless, to ensure participants paid attention to the entire number rather than the last digit alone, each

test block included two trials where the potential successors differed from a correct answer by a multiple of five (e.g. if 23 could be followed by 25, the number 35 might be displayed).

Recall task

The aim of the recall task (Figure 3, below) was to extract a neural or ocular representation of each number following learning. We predicted that these representations would then scale according to action similarity (see below). While some previous studies have successfully isolated representations of relational structure during orthogonal tasks (e.g. Garvert et al., 2017), we decided to use a task in which participants were explicitly asked to recall the previously-learned transitions and tested on this using probe trials. This ensured that the underlying structure (and potential action representation) was engaged with during the scanner session.

As a general overview, during eye-tracking and the fMRI session participants were presented with a continuous series of numbers in the centre of the screen, including both previously seen and new numbers (Figure 1b bottom). Each number was presented for 2s, and was followed by an inter-stimulus interval of mean 3s (or 2.5s during eyetracking; see below) during which a fixation cross was presented.

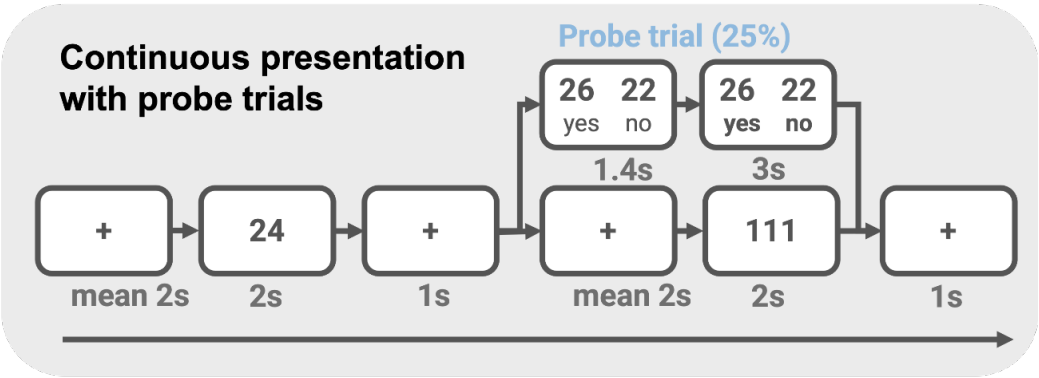

Figure 3. *Participants were presented with individual numbers in pseudorandom order, allowing us to extract the ocular and neural representations for each number in isolation. To ensure attention was paid to the task, probe trials were inserted into the run at random intervals. In these trials, two numbers were presented and participants were asked to identify if these numbers were the correct possible successors of the previous number.*

To keep participants engaged in the task, one quarter of trials were followed by probe trials (presented 1s into the inter-stimulus interval) which queried participants' understanding of the affordance-state bindings. Probe trials consisted of a pair of numbers, which were drawn from a range of two from the previous number. If these two numbers were the two correct successors of the previous number, participants had to press a button to 'accept' them within a short response window (1.4s; 1.2s during eyetracking). Otherwise, they had to reject the option. Half of probe trials showed correct possible successors. After the response interval, either the response selected or a reminder to respond faster was displayed for 3s. Feedback was provided only at the end of runs, as a percentage of correct responses in probe trials, to avoid on-task learning.

Button assignment was counterbalanced across runs, and the ordering of presented probe trial stimuli was randomised, to ensure there was no systematic bias in either motor responses or visual attention. To avoid potential problems of temporal autocorrelation in the MRI scanner, the order of conditions was counterbalanced such that each trial was preceded an equal number of times by trials from different conditions.

Inter-stimulus (ISI) intervals were drawn from a truncated exponential distribution with a mean of 3s in the MRI scanner and 2.5s during eyetracking, respectively. To guarantee that the ISI selection did not influence the results due to temporal autocorrelation, for each run we ensured that the Spearman correlation between ISI length between adjacent trials and model-based trial distance was less than 0.1 via permutation of possible ISI sequences across all models (see 'Analysis: Representational Similarity Analysis' for model details).

In the eyetracker, participants carried out four runs of approximately six minutes each. Each run consisted of 80 trials, of which 16 were probe trials. In the MRI scanner, participants carried out eight runs, each of which lasted seven minutes. As with the eyetracker, each run consisted of 80 trials of which 16 were probe trials. This resulted in a total of 512 trials for analysis.

2.3.2 Data

In this section, I will describe the data acquisition procedures. Both the eyetracking and fMRI data were measured during the recall task, on different days. The eyetracking data was measured at the end of the first day, while fMRI was collected on the second day. The aim was to capture behavioural and neural signatures which could be tested for action representation.

Participants

Sixty healthy German-speaking volunteers participated in the study. All participants were right-handed, with normal or normal-to-corrected vision and reported no history of neurological or psychiatric disorders. All participants provided written informed consent to participate in the study, and were compensated for participation. The study was approved by the local ethics committee at the University of Leipzig, Germany (protocol number 437/22-ek). Three participants did not reach a predetermined performance threshold of 74/128 correct answers in the scanner task, which corresponded to a performance greater than chance in a binomial test at an alpha level of 0.05. These participants were excluded from all analyses. Within the remaining fifty-seven subjects, 29 self-identified as female with an overall mean age of 26.7 (s.d. 4.9). For a further 11 subjects, eyetracking data was incomplete (minimum one full run missing; nine subjects) or inaccurate (all data points further than half of screen width or height from the centre during trials; two subjects) due to issues with binocular calibration resulting from thick glasses, contact lenses or makeup. This resulted in a final sample of forty-six subjects for ocular analyses (21 self-identified female; mean age 25.8, s.d. 4.6).

The sample size was determined without a prior power analysis, due to the difficulties in making an accurate estimate deriving from the small number of previous related studies and the multiple planned analyses. Initially, a high drop-out rate was expected due to a taxing task involving numbers, and therefore it was decided to recruit sixty participants in order to match or exceed the larger sample sizes typical of studies which have been able to detect comparable effects in the entorhinal cortex (Garvert et al., 2023; Nitsch et al., 2024; Viganò et al., 2023).

Eyetracker data acquisition

Participants were seated in front of a monitor leaning on a chin rest and forehead bar, at a distance of 57cm from the screen. We recorded binocular gaze position continuously using an EyeLink 1000 Plus (SR Research), at a sampling rate of 1,000 Hz. To ensure accurate recording, we calibrated the system at the start of each run using the built-in calibration and validation protocols from the EyeLink software using nine fixation points. Validation was accepted when gaze points were less than one degree from points. Accuracy (mean distance from the screen centre) and precision (root mean square distance from mean gaze position) were calculated using a one second window before stimulus onset while viewing a fixation cross.

Precision had a mean of 12.83 pixels and a standard deviation of 7.97 pixels across subjects, while mean accuracy was 17.20 pixels, with a standard deviation of 6.65 pixels.

It may be noted that the task is not perfectly optimised for eyetracking signal: during trials, participants are reading numbers on the screen (Figure 3). While this may have introduced noise, our primary hypothesis regarded the fMRI data, and so we prioritised consistency across the recall task on days one and two.

MRI data acquisition

The MRI data acquisition was designed to maximise signal-to-noise in the entorhinal cortex while allowing whole-brain coverage. To this end, a new sequence was piloted at the MRI scanner using a faster repetition time (1s) while maintaining the voxel size of previous studies which found entorhinal effects (Viganò et al., 2024). The faster sequence allowed multiple volumes to be captured per trial, which should increase the signal-to-noise per condition. Following recommendations from previous work (Nau, 2022), the acquisition box was manually aligned by the experimenter to line up with the long axis of the hippocampus. This has been suggested as a method to acquire clearer signal in the medial temporal lobes.

MRI data were recorded using a 3 Tesla Siemens Magnetom Prisma Fit scanner (Siemens, Erlangen, Germany) with a 32-channel head coil. Following a localiser scan, blood-oxygen-level-dependent (BOLD) contrast was measured for the eight runs of the recall task using T2*-weighted whole-brain gradient-echo planar imaging (GE-EPI) with a multiband acceleration factor of 5. The fMRI sequence had the following parameters: TR = 1000ms; TE = 22ms; voxel size = 2.5mm isotropic; field of view = 204mm; flip angle = 62°; partial fourier = 0.75; bandwidth = 1794 Hz/Px; multi-band acceleration factor = 5; 65 slices interleaved; distance factor = 10%; phase-encoding direction: A -> P. Per run, 415 volumes were acquired.

After every run, fieldmaps were recorded as a way to measure and correct for inhomogeneities in the magnetic field. Fieldmaps were acquired using opposite phase-encoded EPIs, using the same voxel sizes and field of view with TR = 8000ms; TE = 50ms; flip angle = 90°; partial fourier = 0.75; bandwidth = 17.934 Hz/Px. After each scanning session, an anatomical scan was recorded. This was done using a T1-weighted MP2RAGE (TR = 5000ms, TE = 19.6ms, voxel size = 1 mm isotropic). The resulting scan was denoised using the method outlined in (O'Brien et al., 2013)**.**

All stimuli were projected onto a screen of resolution 1024x768 pixels situated above the participant using a mirror attached to the head coil. Behavioural responses were collected using an MRI-compatible button box.

Preprocessing of fMRI data

Preprocessing was carried out using fMRIprep 22.0.1 (see 'Additional Details' for a boilerplate provided by fMRIprep). fMRIprep is a toolbox which follows a standard set of minimal preprocessing steps, designed to ensure reproducibility. The steps include segmentation, head motion estimation and correction, slice timing correction, co-registration to anatomical images, and fieldmap-based (AP-PA) susceptibility distortion correction. All analyses were carried out in volumetric space, using subject space (T1w) for all analyses carried out at the single-subject level. Before running any general linear models (GLMs), volumetric data was smoothed using a Gaussian FWHM filter with a 5mm width.

2.3.3 Analyses

We hypothesised that the entorhinal cortex would represent affordances; that is, neural pattern similarity would scale with the number of shared actions across conditions. To test this, we first aimed to verify that participants were able to carry out the task using their performance in test trials during learning, and probe trials during the recall task. We then tested if participants' eye movements were lateralised according to action possibilities, as an indication that they were engaging with numerical information during the task. To test our main hypothesis, we then ran representational similarity analysis in two entorhinal cortex ROIs (one for each hemisphere), in which we correlated neural pattern similarity with the model-predicted pattern similarity based on action similarities. Finally, whole-brain analyses were run and DeepMReye was used to estimate eye movements during the scanner task and link them to the results in the entorhinal cortex.

In this section, I outline the main analysis methods used. For the sake of readability, additional control analyses are detailed separately in the main results section as they arise.

Analysis of behavioural data

The primary aim of behavioural analysis was to establish that participants were able to carry out the task correctly, indicating that they had learned the action distribution for each state. Performance was measured during the training tasks as the percentage of correct responses in the test phases (see above). On the first day, this corresponded to a total of 720 trials (change level of 50%). Performance on unseen numbers was measured using the two extended test phases after 20 and 40 blocks, which consisted of 48 trials. On the second day, participants were presented with 180 test trials on learned numbers and 24 test trials consisting of previously-unseen numbers.

In the recall task, performance was measured relative to probe trials, of which there were 16 per run (therefore 64 in the eyetracker and 128 in the fMRI scanner; chance level 50%).Participants had slightly longer to reply in the fMRI scanner (1.4s instead of 1.2s). Half of the probe trials consisted of previously-unseen numbers. To compare performance on seen and unseen numbers, a two-sided paired t-test was used with an alpha level of 0.05.

Differences in reaction time in the recall task were measured using repeated measures ANOVA (alpha of 0.05) to ask if state identity predicted the mean reaction time for each state, for numbers followed by a probe trial. This was particularly relevant for the second day, to ensure differences in reaction times (and thus potentially task difficulty) could not explain our model of interest. Post-hoc pairwise comparisons were then carried out using paired t-tests across combinations of states, using a two-tailed test with a Bonferroni-corrected alpha of 0.0083 to reflect six pairwise comparisons.

Finally, questionnaire data was gathered to ensure no participants were aware of the experimental question. Participants were asked if they knew what we were testing, along with a variety of other questions designed to understand the strategy they were using.

Analysis of eyetracking data

The principal aim of the eyetracking analysis was to establish that participants were engaging as expected with the numerical features of the task, which provide the basis for action similarities of dissimilarities. We expected that eye movements would mirror the direction of afforded actions: states which afforded only positive actions would elicit more rightward eye movements than those which afforded only negative actions (Figure 4 right). We tested this by looking for a relative skew in the eye position across the trial window for the two states which allowed only positive or negative actions.

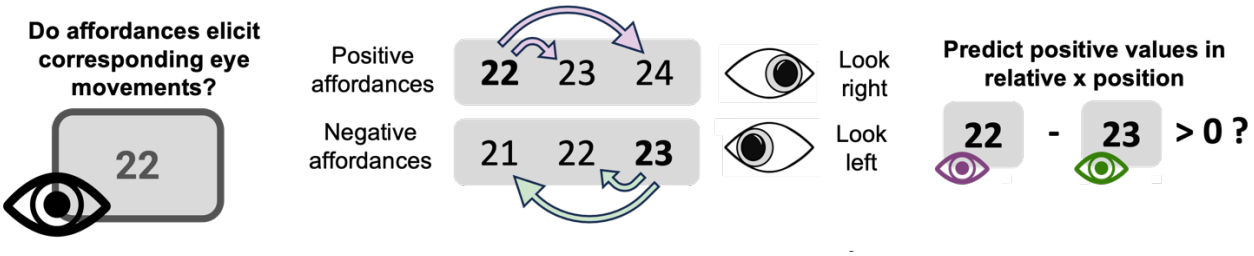

Figure 4*: We predicted that eye movements would how a relative rightwards shift when comparing states which afforded positive actions to states which afforded negative actions.*

The eyetracking data was first cleaned by removing blinks and surrounding data (using automated blink detection carried out by the Eyelink toolbox, and a buffer window of 100 ms) from the time series. As gaze data was recorded binocularly, the mean position in x and y across eyes was taken for each timepoint to obtain a single gaze position for each millisecond. Data was then segmented into trials using a window of -500ms to 2500ms around stimulus onset. Trials of each condition were then grouped together and the median gaze position in x and y was taken for each timepoint. To account for temporal noise, the data was then minimally smoothed with a Gaussian kernel with a standard deviation of 4.25ms (roughly corresponding to a full width half maximum kernel width of 10ms) and a maximum window of 20ms within each condition and subject. This resulted in a single time series in both x and y for each condition, for each subject.

To test for an action-based change in gaze behaviour, we compared the directional shift as a function of different states. State 2 allows only positive operations, while state 3 allows only negative operations. We obtained the relative shift by subtracting fixations for state 3 from state 2 in the x axis, for every time point. If the relative position difference is positive, this means an eye movement further right for states which afford positive actions relative to states which afford negative actions.

In order to carry out group-level analyses, we ran a temporal cluster correction using the MNE package (Gramfort et al., 2013) based on a 1-sample t-test with 10,000 permutations. A two-tailed test against 0 was used with an alpha level of 0.05.

Finally, we compared the results to performance. To do so, for each participant we identified the median change in gaze position across every timepoint within all previously-identified significant clusters. This

gave a single value which reflected how strongly the directional effect was for a given participant within the cluster. A more positive value indicated a stronger effect. This was then correlated with performance in probe trials (Pearson r) in the first day to obtain a measure of how much gaze effects reflected individual performance. An alpha level of 0.05 was used.

Representational similarity analysis: region-of-interest analyses

If the entorhinal cortex represents actions, we predicted that neural pattern similarities would scale with the similarity in action distribution between states. To test this, we used representational similarity analysis (RSA; Kriegeskorte et al., 2008). RSA is a correlational method which compares a matrix of (dis)similarities of neural patterns across different conditions to a matrix of predicted model-based distances. In our case, we would expect that for condition pairs with more similar action possibilities (e.g. both allowing -2 and +2, as for states 1 and 4), we would expect more similar neural patterns.

As a general overview, our approach was to first run a targeted ROI approach in the entorhinal cortex, to test our main hypothesis. We then moved onto a whole brain searchlight approach (see below). As this was not informative due to no clusters surviving statistical testing, we then tested two alternative regions (bilateral medial prefrontal cortex, mPFC, and superior parietal lobe, SPL) as well as the wider medial temporal lobe (MTL) in order to establish the wider distribution of action representation in the brain. Extra controls were carried out, and are detailed in the main results section, in order to establish that the effects we found were not driven by noise. The diagonal was excluded from all analyses (including in the searchlight and subsequent chapter), as for crossnobis distance measures this would be equal to zero and therefore could have driven false positives.

To run RSA, we first ran a first-level GLM on the preprocessed data in order to extract beta values for each condition. The GLM included motion regressors from fmriprep (see above), as well as an intercept. We also included regressors modelling visual information displayed on the screen and motor actions, specifically separate regressors for each type of visual presentation (stimulus presentation, probe trial presentation, response presentation, fixation cross) and for left/right button presses. As conditions, we included the state identity for each presented number, modelled as a stick function. This model was run separately on each of the 8 functional runs in native (subject) space (T1w), to obtain a map of beta values for each condition, run and every subject.

For each ROI, we then calculated the dissimilarity between conditions, in order to generate representational dissimilarity matrices (RDMs) which could be correlated against our model matrices. As we had separate beta maps for each run, we could use cross-validated dissimilarity measures, which identify signal common to the same condition across runs and thereby reduce the effect of run-wise noise. The cross-validated Mahalanobis distance ('crossnobis'; Diedrichsen et al., 2016; Walther et al., 2016) was used, using the residuals from the first-level GLM to estimate the relevant noise covariance matrix for each run. This state-of-the-art method produces a single RDM per subject, in which the distances reflect how consistently dissimilar two neural patterns are across runs. A negative distance would indicate noisy data across runs; a distance of zero would indicate a high similarity; a positive distance would indicate that, across runs, two conditions consistently evoke different neural patterns.

As the main hypothesis concerned a specific region (namely, the entorhinal cortex), a region-of-interest (ROI) approach was used. ROIs were sourced from the Jülich Atlas (Amunts et al., 2020, version 3.0.3) in MNI space, taking the maximum probability maps (MPMs) for each region (right entorhinal cortex, left entorhinal cortex, visual cortex (V1+V2) for control analyses). Maximum probability maps were used to give a conservative estimate of the extent of each region-of-interest, rather than a threshold-based approach which may have accidentally taken into account voxels better identified as hippocampal. ROIs were then mapped into subject space using ANTs and the transformation matrices output by fmriprep. To ensure that the transformation accurately captured grey matter, they were then intersected with a grey matter mask from the fMRIprep output, thresholded at 0.5. The resulting entorhinal ROIs had the following sizes: left entorhinal cortex (mean 111.5 voxels, standard deviation 12.3), right entorhinal cortex (mean 140.1 voxels, standard deviation 22.0).

Using these ROIs, we calculated a neural dissimilarity matrix for the four different conditions in the experiment (each state), using the method outlined above. This was then compared to model matrices (Figure 5) using a tie-corrected spearman rank correlation, resulting in a single correlation value. Partial correlation (using recursion based on tie-corrected spearman rank correlation) was run to account for the effect of other models, which are outlined below.

In order to perform group-level inference, the resulting correlation values for each participant were Fisher z-scored (to approximate normality) and tested against a null hypothesis of mean 0 or less using a one-sample, one-tailed t-test. Directional tests are standard in the RSA literature as negative results are often hard to interpret (see the default implementation in rsatoolbox; Nili et al., 2014; examples can be found in most RSA papers cited in the current work, e.g. Baram et al., 2021; Kaiser et al., 2022), although researchers

have highlighted that one-tailed t-tests in general can enhance false positive rates due to the counterfactual scenario of finding unexpected results in the opposite direction (Lombardi and Hurlbert, 2009). In our case, a negative RSA result would be very difficult to interpret (and reported as non-significant) and so we consider that a one-tailed test is appropriate. However, to ensure clarity for readers I have specified this clearly in the results along with t and p-values. To account for multiple hemispheric comparisons, a Bonferroni-corrected alpha level of 0.025 was used to determine statistical significance in our main affordance model; where a different threshold was used, this has been stated in the main text.

Representational similarity analysis: searchlight

Although our primary hypothesis concerned the entorhinal cortex, ROI analyses can artificially inflate false positives if the effect is widespread, or signal bleeds from a nearby region. Moreover, it is likely that actions are encoded in a variety of cortical regions, such as parietal and prefrontal areas. To test for the specificity of the affordance effect, we repeated the same RSA approach across the whole brain using a searchlight.

Searchlight analysis was run following the procedure outlined in (Kriegeskorte et al., 2006). The logic of this analysis is that representational similarity analysis can be run within spheres to provide a whole-brain map of multivariate responses to different models. In our case, for every voxel in each brain we isolated a sphere of 7.5mm, and used the beta values of the voxels within this area to approximate neural dissimilarities for that voxel.

As before, all analyses were first carried out in subject space. Spheres were generated using rsatoolbox for Python v0.2.0 (Nili et al., 2014; Bosch et al., 2025), with a radius of 3 voxels (7.5mm) and a minimum number of surrounding grey voxels necessary set as 30%. This analysis was run at a whole-brain level.

Within each sphere, we followed the exact same approach as for the ROI-based RSA analysis; that is, neural dissimilarity matrices were calculated using crossnobis distance (with residuals from the first-level GLM for noise covariance calculation), and then compared to model matrices using a tie-corrected spearman correlation prior to z-scoring. We excluded the effect of other models using partial correlation. Fisher z-scored correlation values were replaced in the centre of the searchlight sphere, to generate a whole-brain map of correlation values for each participant in their native space.

In order to run group-level analyses, the resulting maps were then converted back into MNI space using ANTs, and a one sample t-test against 0 was run for every voxel across subjects. To account for the problem of multiple comparisons, TFCE-based cluster correction (Smith & Nichols, 2009) was run within the whole brain. An alpha level of 0.05 was used to determine if a cluster could be considered significant, with no threshold on the number of contiguous voxels to form a cluster.

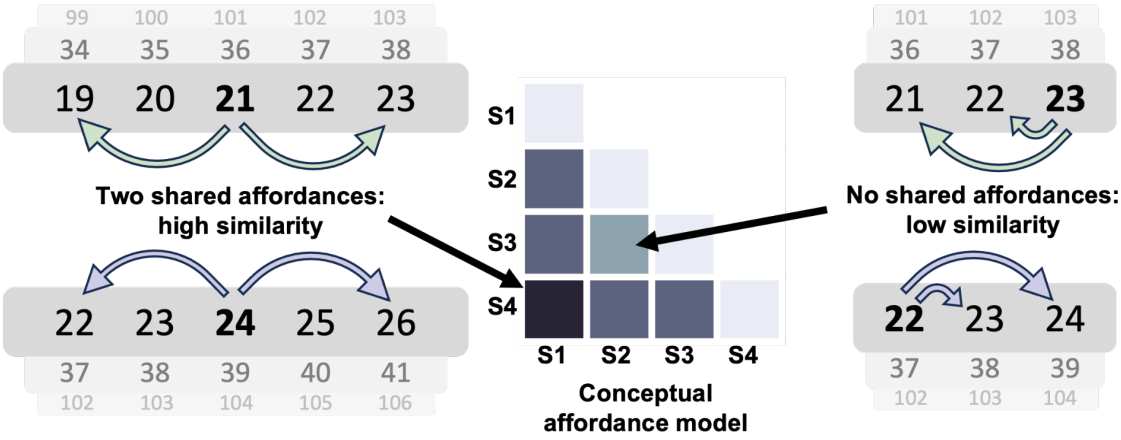

Figure 5*. We created a model of state similarity based on the number of shared actions between states. This was then correlated with neural pattern similarities in the entorhinal cortex in each hemisphere (main analysis), or with searchlight sphere pattern similarities (whole-brain analysis).*

Representational similarity analysis: models of state similarity

Our main model of interest concerned the similarity of states in terms of whether they allowed the same or different actions (Figure 5). To test this, we used a count-based dissimilarity matrix which indicated the number of shared actions between states (subtracted from the maximum, 2). For example, as states 1 and 4 both allow the same two actions (-2, +2) their dissimilarity is 0; states 1 and 2 only have one shared action (+2), so their predicted dissimilarity is 1. This model, therefore, allowed us to test if neural patterns scaled with the number of shared actions between states.

The count-based approach was also used for the two other control models, link distance (the minimum number of steps between two states), and action magnitude (the number of shared actions, independently of the numerical sign). It should be noted that action magnitude matrix is perfectly anticorrelated with a

model based on successor identity (that is, whether you can reach the same state or not). Therefore, excluding this model will also exclude effects linked to the next state identity, analogous to a one-step successor representation.

DeepMReye

Although we did not gather eyetracking data on the second day, the eyetracking results on the first day prompted us to investigate if there might be a connection between the gaze lateralisation effect observed previously and entorhinal representations of action. To gain an estimate of eye position on the second day, the DeepMReye toolbox (Frey, Nau et al., 2021) was used. This toolbox uses a pre-trained convolutional neural network (CNN) to estimate gaze position based on the magnetic resonance signal of the eyeballs.

Following the minimal fMRI preprocessing steps outlined above, gaze positions in x and y (in visual angle) were estimated at a resolution of 10 Hz (ten estimates per TR) using pre-trained model weights based on a task in which participants sequentially fixated different images (Dataset 6).

To gain a single estimate of gaze position within each trial, the median x position per trial was taken within a slightly-extended window of 1s centred around the cluster identified on the first day (to ensure an even number of datapoints per trial, and in line with the standard DeepMReye pipeline of taking the median within a single TR-length window). Within each subject, the median was then calculated across trials of each condition. To compare gaze lateralisation across subjects, a paired t-test was run predicting that the median gaze position would be further right for condition 2 than condition 3 (alpha level of 0.05).

To compare the resulting effects to the MRI data, we firstly compared the extent of gaze lateralisation with the predicted neural differences between conditions 2 and 3. To account for differences in noise levels, each RDM was normalised and the distance between conditions 2 and 3 was extracted. This was then correlated with the difference in gaze position for each subject for these conditions (Pearson correlation, alpha=0.05). Following this, we then checked in an exploratory analysis for a relation between the magnitude of gaze lateralisation and the correlation in the entorhinal cortex with the affordance model (Pearson correlation, alpha=0.05).

2.4 Results

In this section, I will present the results of the experiment and analyses described in the previous section. Firstly, I aimed to establish that participants were able to carry out the task, before using eye movement analyses to test if participants were engaging with the numerical features of the design. Thereafter, we tested our main hypothesis, which predicted that we would find a representation of conceptual affordances in the entorhinal cortex, using the targeted ROI-based RSA approach described in the previous section. The resultant effect was resilient to a variety of control analyses and appeared to be driven by signal-containing voxels rather than noise. On the other hand, it should be noted that no whole-brain effects for the same model (using a searchlight) survived multiple comparison correction, although subsequent ROI-based analyses identified a representation of action in the superior parietal lobe. Finally, I tried to understand if the results in the right entorhinal cortex could be linked to the eye movement lateralisation effect observed on the first day. All results are presented with raw p-values for clarity. A standard statistical threshold of 0.05 is used for significance testing, except where multiple comparison correction is carried out across hemispheres or for multiple comparisons following the logic described above.

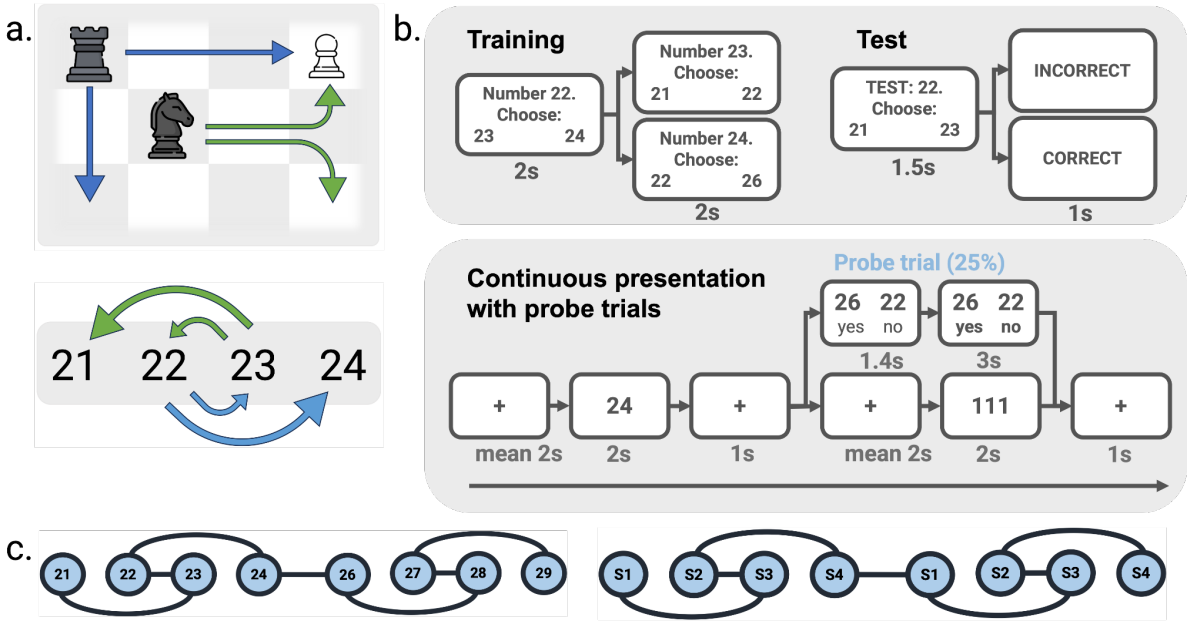

Figure 6: Experimental design. *a. Representations of (conceptual) space may benefit from the integration of action affordances, e.g in a game of chess, where information about how pieces can move is necessary to play the game. We tested this idea using numerical operations. b. Participants learned how numbers may 'transform' into other numbers using an exploration task, whereby they were presented with an individual*

number and were offered a free choice of two possible successors. When a successor was chosen, the next trial offered the successors for the chosen number. In interleaved test blocks, participants indicated which were the 'correct' successors for each number. In the eyetracker and fMRI scanner, individual numbers were presented in a pseudorandom order. To ensure attention on the learned actions, one quarter (16) of trials were followed by probe trials in which participants indicated if two presented numbers were the correct successors. c. Unbeknownst to participants, number-action bindings formed part of a repeating sequence arranged in graph 'modules' of four numbers. This can be seen as a set of four states with the same action affordances. As the modules repeated every five numbers, the structure allowed easy inference of action possibilities for any new numbers. Chess images adapted from flaticon.com.

2.4.1 Learning and generalisation of state-action associations

We predicted that if the entorhinal cortex represents transition information in a conceptual space, then neural pattern similarities would scale with the action possibilities afforded by different states. Therefore, we first aimed to teach participants a repeating graph consisting of states and actions along a number line: participants learned that certain numbers permitted transitions to other numbers (according to a predetermined graph structure, see Figure 6c) by free exploration along the graph using button presses. For instance, state 1, which could be instantiated by different numbers, afforded +2 or -2, while state 2 afforded +1 or +2. Knowledge of the graph was tested using blocked test trials presenting two candidate successors, of which only one corresponded to a possible transition and had to be chosen. In two subsequent sessions using the eyetracker and in the MRI scanner, respectively, participants were presented with individual numbers in a pseudorandom order (Figure 6b top). In these sessions their knowledge of the graph was probed using probe trials showing two numbers, to which participants had to respond if both were the correct successors for the previously-presented number or not (Figure 6b bottom) .

Participants were trained on two consecutive days to learn how specific numbers could 'transform' into other numbers. By the end of the first day, participants were able to reliably select the correct transitions for different numbers (accuracy: 84.7%; chance level 50%, std. 7.1%, last ten blocks; figure 7a). Moreover, they were able to apply the learned transformations to new numbers: after training on the second day, they reached a performance level of 88.8% on unseen numbers (std. 11.3%). The application of these operations to hitherto-unseen numbers indicates that participants understood that numbers and transitions were factorised.

In the recall task on the first day (during eyetracking), performance was lower than in the previous task (accuracy: mean 68.1%, std. 15.9% across day 1; missed responses: mean 8.4%, std. 6.4%), but nonetheless improved as familiarity with the task increased to 71.5% in the fourth and final block (std. 21.6%; difference between fourth and first block T=218.5, p=0.006, Wilcoxon signed rank test). Participants were better at answering correctly for seen than unseen numbers (seen: mean 70.8%; unseen: mean 65.8%; t(56)=2.78; p=0.008).

Knowledge of the structure persisted to the second day, suggesting a long-term encoding of the graph information. In the MRI scanner, performance was at a mean accuracy of 82.7% (std. 12.5%; missed responses: mean 7.8%, std. 6.3%). Moreover, the differences between seen and unseen numbers diminished to a non-significant level, suggesting that longer-term learning was dominated by the abstract structure rather than number-specific action bindings (no significant difference between seen and unseen numbers: seen: mean 83.2%, std. 12.8%; unseen: mean 82.2%, std. 12.3%; t(56) = 1.56, p=0.13; see Figure 5c).

A stimulus-independent encoding of the graph structure was supported by state-specific reaction time differences during probe trials. Reaction time data in the recall task on the second day showed a difference between response times for different states: state identity predicts reaction times ($F(3, 168)=9.77$, $p<0.0001$). Post-hoc pairwise comparisons show a significant difference between states 1 and 2 ($t(56)=4.44$, $p<0.0001$), states 1 and 3 ($t(56=3.45$, $p=0.0011$), states 2 and 4 ($t(56)=-4.09$, $p=0.0001$) and states 3 and 4 ($t(56)=-3.38$, $p=0.0014$). There were no significant differences between states 1 and 4 ($t(56)=0.17$, $p=89$) and states 2 and 3 ($df(56)=-0.23$, $p=0.82$).

In short, participants were able to learn the numerical operations afforded by different numbers, and apply this knowledge to new numbers. This indicated knowledge of a repeating structure consisting of state-action bindings.

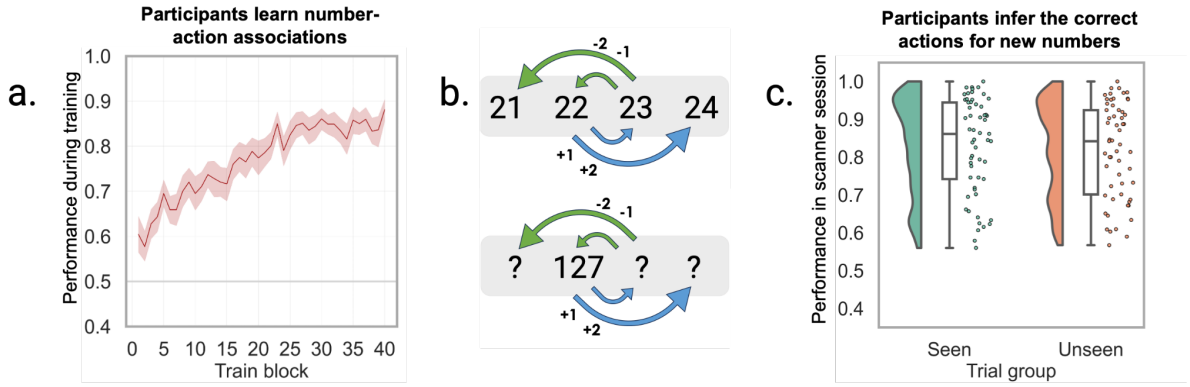

***Figure 7: Behavioural performance**. a. Participants learned the numbers and successors well, achieving a performance of 84.7% by the end of the first day. b. The repeating structure should allow participants to infer the correct successors for previously-unseen numbers. c. In the recall task in the fMRI scanner, participants were able to correctly identify the successors when presented with unexpected probe trials (mean performance 82.7%). There was no difference in performance between previously-seen numbers and new numbers in the scanner task (p=0.126).*

2.4.2 Eye movement direction reflects the sign of conceptual affordances

Prior work has suggested that when participants engage with numerical information, eye movements skew left or rightwards based on numerical magnitude or the direction of arithmetic (Loetscher et al. 2008, 2010; Viganò et al., 2024; see the discussion in this chapter for a discussion of the origin of this lateralisation). In our task, actions similarities relied upon participants engaging with the numerical features of the task, rather than learning purely via visual or linguistic associations. We reasoned that if participants were processing numerical information, eye movements would skew left or right based on the type of actions allowed. Based on previous studies, we expected eye movements to skew more rightwards for states that allow only positive actions (e.g. state 2) relative to states that allow only negative actions (e.g. state 3).

We compared the relative gaze position difference for these states for each timepoint in a [-0.5,2.5] window around stimulus onset and tested for each timepoint in the trial if movement shift was different to 0 along the horizontal axis, before running cluster correction to account for multiple comparisons. We found one cluster where this was the case ($t(45)=5.49$, $p=0.0003$, cluster window of 922 to 1739 ms after onset; Figure 8b). Finally, we show that ocular responses were skewed more strongly left-or-rightwards in better-performing participants ($r=0.493$, $p=0.0005$, $n=46$; Figure 8d). Therefore, participants who expressed numerical information more through eye movement were more successful in solving the task.

In sum, the behavioural results show that participants had learned and generalised the affordances across different instances (e.g. new numbers) of the states and their reflection in eye movements is congruent with an engagement with the numerical features of the task. Thus, next we approached our main question of whether the entorhinal cortex would represent states in terms of the actions they afford.

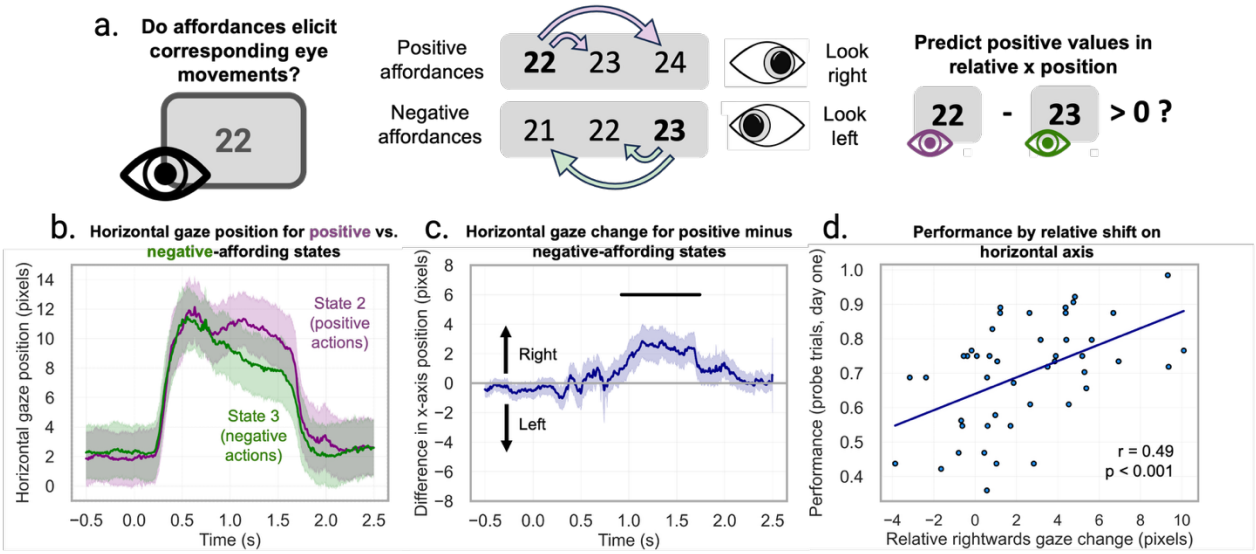

Figure 8: Ocular representation of conceptual affordances. *a. We predicted that, if participants were engaging with magnitude information, eye movements would skew further right for numbers which afford positive actions, relative to numbers which afford negative actions. b. Relative gaze positions for states 2 and 3. When the mean x position of state 3 (which is mapped to number 23 in the example) was subtracted from the mean position for state 2 (mapped to number 22), we find a positive value, indicating a more positive shift for more positive actions. Horizontal bars indicate significance. c. The extent of the effect (median position change within the cluster window) correlated with performance in the probe trials of the recall task during eye-tracking.*

2.4.3 Entorhinal pattern similarity reflects affordances of states

The main prediction was that the entorhinal cortex would represent conceptual affordances; that is, neural responses would be more similar when states allowed similar actions. To test this, we used representational

similarity analysis to compare neural pattern similarity across the states to a model RDM indicating the number of shared possible actions between states. Indeed, we found a representation of conceptual affordances in the right entorhinal cortex (right: t(56)=2.13, p=0.019; left: t(56)=-0.49, p=0.69; one-tailed one-sample t-test against 0; see methods). We further assured that this result is not driven by other effects related to affordance magnitude (the unsigned magnitude of afforded actions) or link distance, by including these alternative model RDMs (see Methods; see model RDMs in Figure 9c) as covariates in a partial correlation and replicating the finding (right: t(56)=2.29, p=0.013; left: t(56)=-0.60, p=0.73; Figure 4a-c).

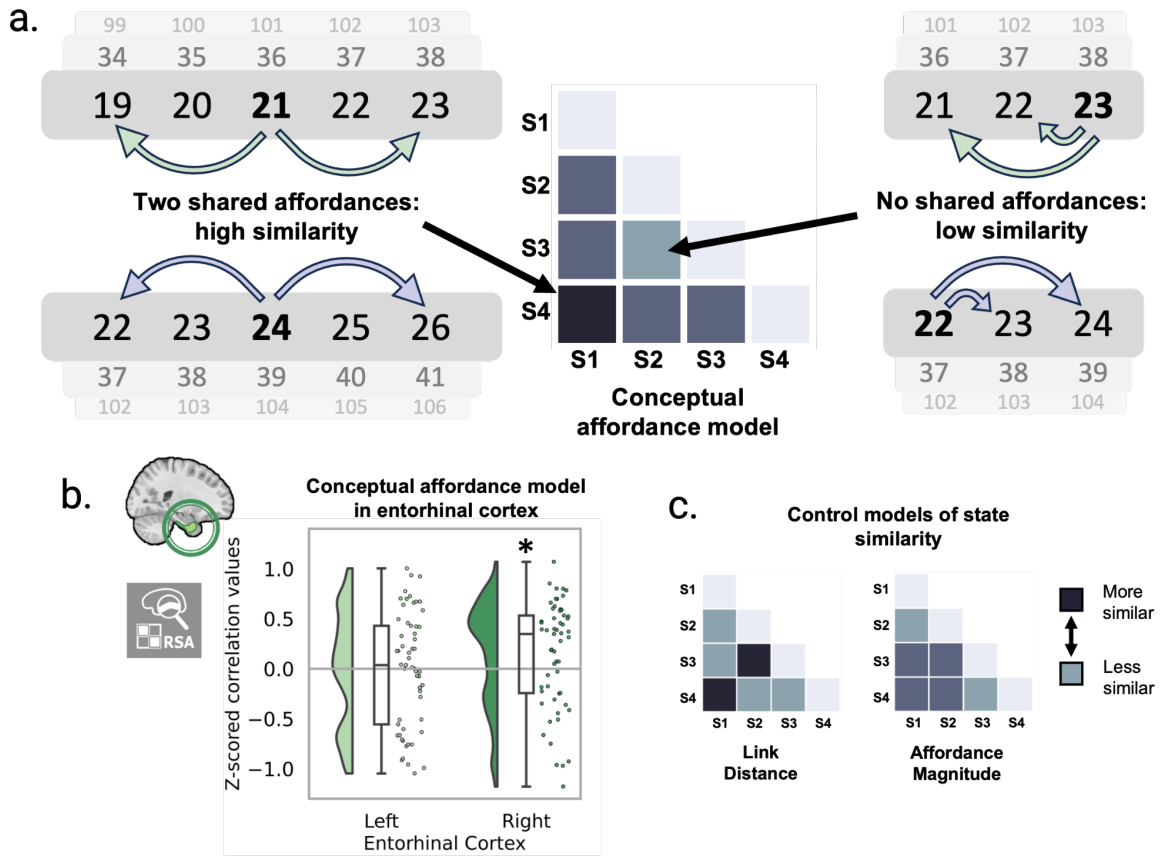

Figure 9: Neural representation of conceptual affordances*. a. We used representational similarity analysis to compare neural pattern similarity across states in the entorhinal cortex with a model of state similarity based on the number of shared affordances. b. Neural pattern dissimilarity matrices (neural RDMs) were calculated in an a prior region-of-interest (ROI) using cross-validated Mahalanobis distance. Neural patterns in the right entorhinal cortex were correlated with the conceptual affordances RDM. c. Control models of state similarity. Asterisks indicate significance.*

Although this result aligned with our predictions, it is important to note that the entorhinal cortex is typically a highly noisy region. Therefore, I wanted to ensure that the observed action representation could not simply be the result of random noise. I therefore aimed to test this both at the multivariate level by looking at the level of noise in the neural representational similarity matrices, as well as at the single-voxel level by selecting only condition-responsive voxels.

Test of noise in cross-validated RSA

As crossnobis is a cross-validated distance measure, we can also use the resulting similarity scores to assess how much a given region is consistently responsive to the conditions across runs. Namely, a score greater than 0 indicates a consistent response (Diedrichsen et al., 2016; Walther et al., 2016). To test this, for each region we calculated the mean distance within the neural RDM for each ROI and tested against 0 using a Wilcoxon signed rank test. Note that this is a conservative test, as we predict a distance of 0 for some state combinations in the affordance model.

Both regions showed mean distance values greater than 0 (Wilcoxon signed rank test, $p<0.0001$ for both ROIs), indicating the results were driven by signal (that is, consistent voxel patterns across runs) rather than noise.

It is worth noting, furthermore, that if only subjects with the mean distances greater than 0 are selected, the affordance RSA effect in the right entorhinal cortex remains ($n=35$, $t(34)=3.15$; $p=0.0046$; one-tailed, one-sample t-test against 0), whereas in the left entorhinal cortex this is not the case ($n=40$, $t(39)=-0.0090$; $p=0.50$; see Figure 10). This further indicates that for subjects with consistent neural pattern responses between runs for each condition (i.e. BOLD responses explained by conditions of our GLM, not condition-irrelevant noise), the right entorhinal cortex specifically represents conceptual affordances.

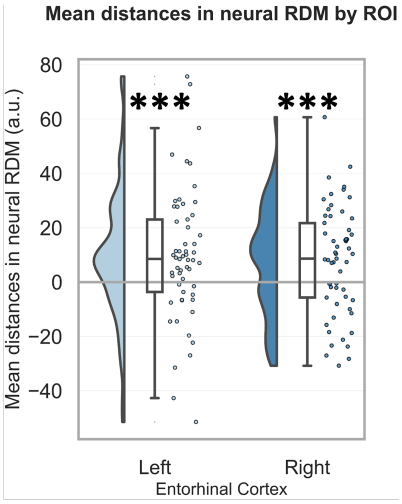

Figure 10*: Cross-validated distance measures produce distance values greater than 0 if there are similar voxel patterns across runs. This allows testing for noise within specific ROIs. In both ROI used, we find mean distance values across all cells of the RDM matrix were greater than 0 (Wilcoxon signed-rank test, p<0.0001).*

Reliability-based voxel selection

As mean cross-validated neural distances were significantly greater than 0 at a group level, it may be concluded that the effect is not driven by noise at the multivariate level. Nonetheless, I was interested to establish that noise in individual voxels could not be a source of Type 1 error.

To ensure that the main effect was not driven by noise at the voxel level, we implemented a modified version of reliability-based voxel selection (Tarhan & Konkle, 2020), which is often used in decoding to reduce the predominance of noisy voxels. The logic is elegant: if a voxel is meaningfully responsive to the experimental conditions, it should respond the same way to each condition in two halves of the dataset. Therefore, one can divide the data in two halves and, for each voxel, correlate a list of beta values for conditions across each half of the data. This provides a correlation value for each voxel; the higher the correlation, the more condition-selective the voxel.

The initial method was designed for condition-rich designs, and the authors advise caution in multivariate analyses due to potential imbalances. As the trials were balanced across conditions, we considered that the

method was appropriate for RSA. Nonetheless, as the design had only four conditions, it was necessary to adapt the method to be sure that it was robust.

To do this, we considered a permutation-based approach. Rather than simply correlating betas in averaged halves of the data across conditions (as with the original method), which would involve a 4-by-4 correlation, we chose to (1) use Euclidean distance between vectors of beta values, which is more robust than correlation for small vectors, and (2) shuffle the data across conditions for a given voxel (1000 times), and re-calculate resultant distances to obtain an estimate of how likely a given Euclidean distance was based on the distribution of possible distances. Using this approach, we could assign a probability value to each voxel which described how much of the distribution was less than the actual, non-shuffled distance measure. A low probability value indicates a consistent response for a given voxel for each condition across runs.

The resultant probability maps were intersected with entorhinal ROIs to generate smaller ROIs within which voxels were more likely to be condition-responsive. Checks on simulated data confirmed that this procedure did not increase the rate of false positives. Nonetheless, this analysis selects voxels where conditions are represented similarly across runs, so it should be noted that results using this method should be interpreted as "[non] significant given that the brain region of interest is differentially responsive to the experimental conditions".

The resulting probability values were quite high across the whole brain, which could be driven by the method itself or a low selectivity to conditions at the single voxel level. Therefore, we chose a liberal threshold of $p<0.2$ for the intersected ROIs (resulting in small ROIs: right entorhinal cortex mean voxels 31.2, std. 17.4; left entorhinal cortex mean voxels 22.3, std. 10.4). Despite the much smaller regions-of-interest, we still found a significant affordance effect in the right entorhinal cortex ($t(56)=2.30$; $p=0.0126$; one-tailed one-sample t-test against 0), suggesting the main finding is not the result of variations in voxel-level noise captured within anatomical regions-of-interest (left entorhinal cortex: affordance: $t(56)=0.75$; $p=0.227$; see Figure 11).

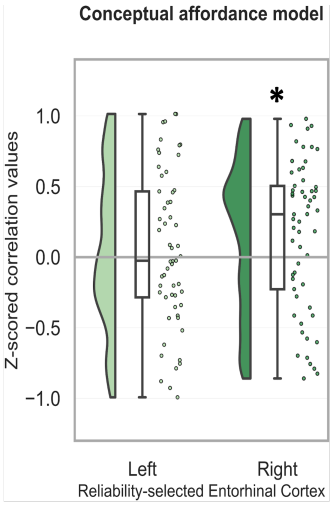

Figure 11*: Using a variant of reliability-based voxel selection, we created smaller regions-of-interest in which neural responses were more consistent, for each voxel, across all eight runs. Within these reduced ROIs, the main effects found in the primary analysis were still present, further confirming that our results are driven by signal rather than noise.*

Comparison of affordance effect in right entorhinal cortex to noise ceiling

Given the above considerations, the effect in the right entorhinal cortex appears to be driven by condition-responsive voxels, and multivariate analyses capture meaningful condition-wise differences. Nonetheless, it should be noted that the result in itself is not particularly strong – could another model explain our neural dissimilarity matrices better?

One way to quantify this is by comparison to the lower bound of a noise ceiling, which quantifies the best possible fit of a model across participants given inter-participant variability – that is, the maximum fit possible given noise. If the fit of a given model is below this noise ceiling, it would indicate that another model may fit the given data better.

The noise ceiling can be calculated using a cross-validated method which compares each RDM to a group average of the other RDMs. I used inbuilt functions from rsatoolbox to quantify this (Nili et al., 2014; Schütt et al., 2023; Bosch et al., 2025) and tested the difference between the 'optimal' model correlations (lower noise ceiling) and the affordance model. The affordance model was not significantly below the noise ceiling ($t(56)=0.69$, $p=0.246$, one-tailed paired t-test). Therefore, there is not good evidence to suggest that

our affordance model is worse than an optimal model of the data (given noise), assuming the noise ceiling correctly approximates an optimal model. Thus, another model would not explain our results significantly better than the affordance model given inter-subject variation.

Affordance effect in right entorhinal cortex is not driven by task difficulty

The above control analyses appear to confirm that the neural pattern similarities are indeed well-represented by our hypothesised model. Nonetheless, we observed a difference in reaction time as a function of condition. This may indicate meaningful differences in task difficulty between conditions, insofar as this may be captured by reaction time differences. To ensure that this was not the case, we ran a control analysis in which we regressed out a model RDM based on average reaction time by state (for each participant) using a partial correlation. The effect in the entorhinal cortex survived this correction ($t(56)=2.079$, $p=0.0211$; one-tailed one-sample t test against 0), indicating that our result is unlikely to be driven by task difficulty.

2.4.4 Whole brain results

As noted in the Methods section, ROI analyses can be deceiving: signal could derive from another neighbouring region, or represent part of a wider system. In our case, for example, the affordance signal could derive from another region of the medial temporal lobes. More generally, characterising the whole-brain response to affordances would help us understand how specific action representation was to the right entorhinal cortex.

To test this, we first ran a whole-brain searchlight as described in Methods. For each participant, each voxel was taken as the centre of a sphere, which was used to calculate a neural RDM for each voxel following the same procedure as within the entorhinal ROIs. These were then compared to the affordance model RDM using the same procedure as for ROI-based RSA, resulting in whole-brain maps of (Fisher z-scored) correlation values. After mapping these to MNI space, we then tested each voxel against 0 using a one-sided t-test, and ran threshold-free cluster enhancement (TFCE) to account for multiple comparisons. This resulted in a corrected statistical map of voxelwise responses to the affordance model.

No significant results emerged from the searchlight (all $p>0.1$). There are various possible reasons for this, such as distributed activation patterns not being captured within the searchlight sphere (although this is unlikely to be the case within the entorhinal cortex due to the small ROI), not forming coherent clusters after transformation to MNI space, or inaccurate estimation of noise in searchlight spheres resulting in noisier neural distance measures. Given that our entorhinal result was not particularly strong, although robust (i.e. $p>0.01$), the effect may simply be too weak to emerge at whole brain level.

However, we were nonetheless interested to understand better the distribution of action representations across the wider brain. There is good reason to think that action may be widely distributed throughout the cortex: evidence has shown that the posterior parietal cortex represents numerosity and a variety of low-dimensional spaces (Summerfield et al., 2020), as well as action and route planning (Nitz et al., 2014). The prefrontal cortex, meanwhile, is often associated with decision making and action choice (Miller and Cohen, 2001). Finally, we were interested in understanding how specific the action representation was to the entorhinal cortex – our effect could be part of a wider hippocampal system, or localised to the entorhinal cortex itself.

Therefore, we designed ROIs based on these regions. Following the same approach as before, the ROIs were based on the intersection of ROIs drawn from the Juelich atlas maximum probability maps and a thresholded grey matter mask for each subject. The parietal ROI included all regions identified as part of the superior parietal lobe (SPL), the medial temporal lobe (MTL) ROI included the hippocampus and subfields, parahippocampal gyrus and entorhinal cortex (both bilaterally and right-lateralised), and the frontal ROI (hereon mPFC) included medial frontal areas categorised as posterior and superior anterior cingulate cortex. As before, we then tested this by correlating neural RDMs in each region to the affordance model. To account for the multiple comparisons across four ROIs involved here with no particular hypothesis, I use a more stringent alpha threshold of 0.0125 for significance, although it should be noted that none of the following results would be considered significant if the threshold were less strict.

Firstly, we found that the medial temporal ROI did not significantly correlate with the affordance model, either bilaterally ($t(56)=0.66$, $p=0.255$) or when looking exclusively at the right hemisphere ($t(56)=0.51$, $p=0.3071$). This suggests that the right entorhinal effect may be specific to the entorhinal cortex, rather than a wider distributed effect across the whole MTL. Future analyses may aim to test the different areas of the hippocampal formation, although it should be noted that the large voxel size will limit particularly fine-grained analyses.

In bilateral mPFC, we found no evidence for a representation of affordances ($t(56)=1.32$, $p=0.0958$). On the other hand, the SPL showed a very strong effect of the affordance model ($t(56)=4.10$, $p<0.0001$). This is in line with the previously-mentioned evidence that the parietal cortex plays a role in action and route planning, as well as numerical cognition. Nonetheless, it should be noted that this is likely a quite distributed effect, as it did not emerge in the searchlight analysis.

2.4.5 Gaze lateralisation during the scanner session

Finally, we also tested for an affordance representation in eye movements during the FMRI session on day 2, corresponding to our finding on day 1. Since we did not perform eye-tracking in fMRI, we used DeepMReye, a toolbox based on a convolutional neural network (Frey and Nau et al., 2021) to predict gaze position from functional MRI data at a sub-TR resolution. We tested whether participants would look further right for positive- as compared to negative-affording states in the previously-identified cluster window (see Methods) and again find that participants' estimated gaze position tended to be further right for positive-affordance numbers than for negative-affording numbers (paired t-test; $t(56)=2.0279$, $p=0.024$; Figure 12a and b). Using this fMRI-based estimate, the correlation with performance no longer reached significance ($r(55)=0.194$, $p=0.1482$).

We would expect that participants with more pronounced gaze differences would also show greater neural distances in the entorhinal cortex between conditions eliciting lateralised eye movements. Regarding the relation to the entorhinal representation, we found that the (normalised) neural distances between states 2 and 3 correlated with the horizontal distance between median gaze positions ($r=0.263$, $p=0.0484$, $n=57$; see Figure 12c). Accordingly, we also observed a non-significant trend in the correlation between the strength of the entorhinal affordance representation to the magnitude of the gaze lateralisation ($r=0.231$, $p=0.0844$; $n=57$). These may indicate a relation between entorhinal representations of conceptual information and gaze behaviour.

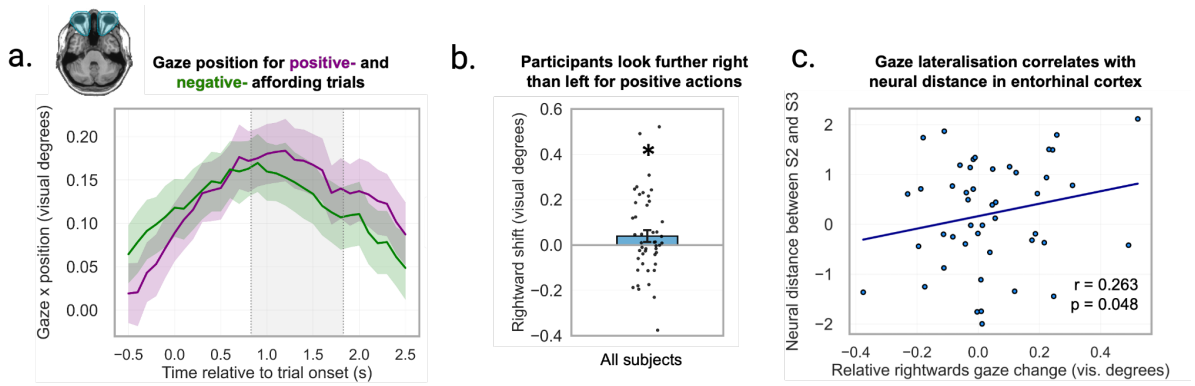

Figure 12: DeepMReye-based gaze lateralisation*. a. Predicted gaze positions during the functional MRI scan were extracted using DeepMReye. We expected that the median gaze position would be further right for positive-affording numbers than for negative-affording numbers during the cluster window identified using eyetracker data from the previous day. Brain image from Frey et al., 2021. b. The median x position within the cluster window was further right for positive-affording trials (state 2) than negative-affording trials (state 3). c. The strength of the gaze lateralisation effect correlated with the predicted neural distances in the right entorhinal cortex (after normalisation across the RDM) for states 2 and 3.*

Given that we also found a representation of affordances in the parietal cortex, we asked whether the strength of the affordance effect in SPL was also related to the extent of gaze lateralisation. Curiously, given the well-established parietal role in attention and eye movement, and the strength of the affordance representation, this did not appear to be the case, either for a correlation between gaze lateralisation and the distance between states 2 and 3 (see above; r=0.103, p=0.446, n=57) or for a correlation between the strength of the affordance effect and gaze lateralisation (r=0.147, p=0.277, n=57).

Given the findings linking the right entorhinal cortex to gaze behaviour, we wondered if the right entorhinal cortex results might be at least partially driven by visual input or retinotopic mapping, rather than the conceptual features of our task. Indeed, mapping of visual space has previously been demonstrated in the entorhinal cortex (Killian and Buffalo, 2018) and the entorhinal cortex has been proposed as a peak of the visual hierarchy (Fellemen and Van Essen, 1991). In this case, we would expect that more similar gaze positions would elicit more similar neural activity. Therefore, we created an alternative RDM based on gaze position in 2D space. This model used the Euclidean distance between the median x and y position for each condition, as predicted by DeepMReye, as a measure of dissimilarity. We used the same time window and method as for the previous analysis (for details, see Methods). We validated this model by showing that it was represented in a visual cortex ROI (t(56)=2.119; p=0.0193). Nonetheless, this model was not

represented by the right entorhinal cortex (t(56)=-0.771, p=0.778), and the affordance effect persisted when this visual model was excluded in a partial correlation (t(56)=1.882, p=0.0369, see Figure 13). This suggests that while our affordance model is connected to decoded gaze behaviour, it seems as though it cannot be explained with reference to eye movement alone.

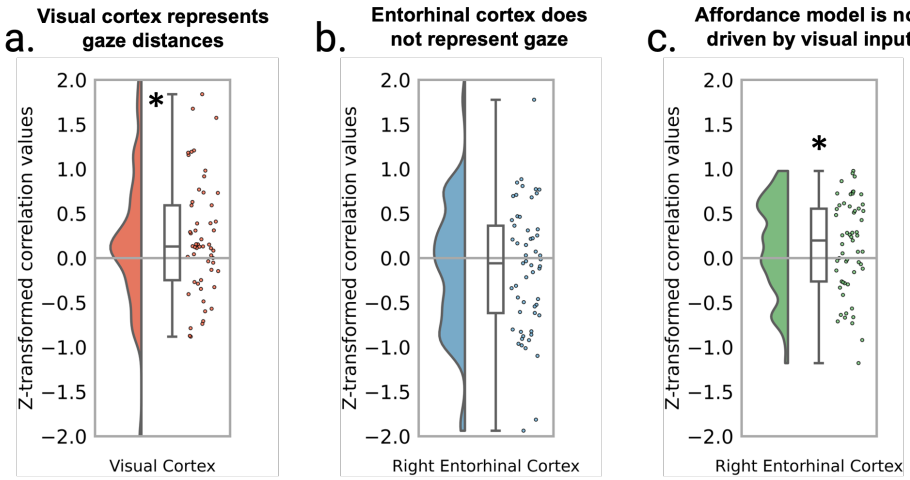

Figure 13: Controls for visual effects in right entorhinal cortex*. a and b. A model of gaze similarity correlates with neural pattern similarities in the visual cortex, but not the right entorhinal cortex. c. The effect in the right entorhinal cortex survived when this effect was controlled for in a partial correlation.*

In sum, we find that conceptual affordances are represented in the entorhinal cortex and indirectly reflected in gaze behaviour. This entorhinal affordance representation was not driven by alternative properties such as link distance or magnitude, or by general gaze behaviour, and is robust with respect to selected participants, voxels, and noise ceiling. Moreover, while searchlight-based whole-brain analyses were inconclusive, a secondary analyses revealed that the superior parietal lobe also represents conceptual affordances.

2.5 Discussion

The hippocampal formation has a demonstrated role in mapping out relational knowledge (O'Keefe & Nadel, 1978; Behrens et al., 2018; Bellmund et al., 2018; Bottini & Doeller, 2020) with potential implications for general cognitive performance (Theves, 2025; Tenderra & Theves, 2025). A functional distinction has been proposed between the hippocampus and the entorhinal cortex, which has been proposed as a neural substrate for the generalisation of relational structures across different sensory stimuli (Behrens et al., 2018). In the absence of idiothetic signals, however, it is unclear how the brain 'navigates' these memory representations - to play a game of chess we need to know both the layout of the pieces and how they move. Prior work has demonstrated that the entorhinal cortex represents action-relevant information during spatial navigation, and models of the hippocampal formation often assume a fixed 'action' input to move between states (Stachenfeld et al., 2017, Whittington et al., 2020, Whittington et al., 2022), an abstraction of the velocity input used in models of spatial navigation (Samsonovich, A., & McNaughton, 1997, McNaughton et al., 2006; Fuhs and Tourtsky, 2006; Guanella et al., 2007; Burak and Fiete, 2009, Banino et al., 2018; Khona and Fiete, 2022, Sorscher et al., 2023; Chandra et al., 2025) with no clear neuroanatomical substrate. Here we provide evidence that the human entorhinal cortex does represent actions in a conceptual space, and that these representations are linked to gaze behaviour. As with prior reports, we note the similarity between the idea of possible action representation and Gibson's theory of affordances (Gibson, 1979; Behrens et al., 2018; Summerfield, 2022)

In particular, we used a number line as base structure to which states with different conceptual affordances were associated, thereby allowing us to distinguish action representations from other features of a state space (e.g. state identity, link distance). We find that the entorhinal cortex represents the actions 'afforded' by a given state. In line with previous work showing lateralised oculomotor response patterns based on numerical sign or magnitude (Dehaene et al., 1993; Hartmann, 2022; Viganò et al., 2023, 2024), conceptual affordances were also tracked by eye movements: participants looked further rightwards when afforded actions were positive relative to states with negative action affordances. There are multiple possible reasons for this, including instantiation of a mental number line (Restle et al., 1970). This could in theory be influenced by other factors such as reading direction (Dehaene et al., 1993 for an early discussion of SNARC reversal in right-left reading cultures, and also Shaki et al., 2009), but our German-speaking participants would be expected to have a left-right number line. Other possible explanations could come from hemispheric dominance (Nuñez and Fias, 2017) or learned motor associations which structure how we handle upper-lower comparisons (Walsh, 2003), as seen in tasks such as finger counting (Fischer, 2008).

In general, however, we take this result as evidence of engagement with numerical information, which has been previously shown to elicit responses in the medial temporal lobes (Kutter et al., 2022). We also show that the gaze lateralisation effect during eyetracking is positively correlated with task performance, possibly indicating allocation of attention (Giari et al., 2023). A potential functional role of eye movements might also be indicated by their relation to entorhinal affordance representations. Our results would go in line with findings that visual grid cells are modulated by recognition memory (non-human primates: Killian et al., 2012; humans: Graichen et al., 2024) and may extend the proposed link between attention and spatial navigation (Nau, Julian et al., 2018) to conceptual spaces (Giari et al. 2023).

More generally, our work fits in with a growing body of evidence that entorhinal representations are tied to learned behaviour. Previous fMRI studies have shown that neural representations in the entorhinal cortex and hippocampus are influenced by task demands or the relevance of dimensions and goals (Theves et al., 2020; Crivelli-Decker et al., 2023; Sweigart et al., 2023). For instance, when learning new concepts, the hippocampus organized stimuli according to the category-defining dimensions and not according to all perceptual or mnemonically-relevant dimensions (Theves et al., 2020). Blind humans exhibit a four-directional symmetry instead of the typical hexadirectional symmetry likely elicited by grid cell populations, suggesting that entorhinal representations may reflect long-term behavioural patterns (Sigismondi et al., 2023). Similarly, rodent and human work has shown a distortion in grid signal as a function of goal placement (Boccara et al., 2019; Butler et al., 2019; Viganò et al., 2023). Moreover, rodent intracranial work has shown that grid cells display realignment toward landmarks, which appears to reflect future behaviour. These distortions are too slow to explain one-shot learning (the authors invoke a mediating plasticity mechanism as a potential solution), but nonetheless suggest a coupling of entorhinal representations with behaviour (Wen et al., 2024; for landmark alignment in a mental simulation task see Hwang et al., 2023; Neupane et al., 2024). Our findings generally support the intuition that entorhinal representations are influenced by possible actions (for a computational model of grid cells based on action constraints, see Dorrell et al., 2022).

Theoretically, there are multiple possible neural substrates for processing of actions in the entorhinal-hippocampal system. In spatial navigation, action is integrated into the grid cell network as velocity input, which may be derived from idiothetic signals or internally generated (Burak & Fiete, 2009; Khona & Fiete, 2022; Dong & Fiete, 2024) and can be described by specialised cell types in the entorhinal cortex (Kropff et al., 2015; Mallory et al., 2021; Munn et al., 2020). Intriguingly, Mallory and colleagues (2021) identify entorhinal cells which conjunctively code for rodent eye and body position, reminiscent of saccade direction cells in head-fixed primates (Killian et al., 2015; Killian & Buffalo, 2018). Beyond spatial navigation, it

has been suggested (Whittington et al., 2020, 2022) that action computations can be carried out through vector cells such as object-vector cells (Høydal et al., 2019) and goal-vector cells (Sarel et al., 2017). In visuospatial tasks in primates, we would suggest that the neural equivalent could be cells with responses similar to saccade direction cells (Killian et al., 2015), but which would fire for both a specific saccade distance and direction.

Based on our fMRI data, it is difficult to draw a direct link between action representation and specific neural populations. Nonetheless, in light of the role for vision in primates and work suggesting a role of eye movements in 'simulating' conceptual navigation (Killian & Buffalo, 2018; Nau et al., 2018; Staudigl et al., 2018; Giari et al., 2023; Graichen et al., 2024; Viganò et al., 2023, 2024), it is conceivable that the observed eye movements reflect an embodiment of velocity input which can enable efficient updating of positional representations in the entorhinal cortex. Although recent work has proposed that velocity input can be stored and retrieved from the hippocampal formation without the involvement of other systems (Chandra et al., 2025), eye movement could provide a useful constraint for generating two-dimensional movement vectors in noisy, high-dimensional domains with no concrete referent. Elucidating the nature of this velocity signal will require further studies specifically aiming to disentangle embodied from disembodied representations of abstract movement.

In terms of the wider brain network for action representation, we find that the superior parietal lobe (part of the posterior parietal cortex) also represents action possibilities. This is in line with prior work in primates demonstrating a role for the posterior parietal cortex in route and action planning (Nitz, 2014; see also Whitlock et al., 2012 and Whitlock, 2014 for a discussion of the different roles of parietal and entorhinal neural representations in rodents in action representation), as well as a wide array of studies finding parietal representations of magnitude (Summerfield et al., 2020) or numerosity (Piazza and Izard, 2009) in humans and non-human primates. One surprise, however, was that the parietal representation of action was not correlated with the strength of gaze lateralisation, despite previous work in non-human primates having identified regions of the parietal cortex as instrumental for handling oculomotor planning (e.g. Crowe et al., 2004) and a possible neural substrate for lateralised magnitude effects (Walsh, 2003) and theoretical work suggesting that the parietal cortex processes self-motion information prior to projection to the medial temporal lobes (Bicanski and Burgess, 2018). This may be due to a lack of parietal involvement but may also result from inaccurate eye movement estimation from DeepMReye, a lack of variance in correlation values, or simply that our model does not capture the way parietal regions handle eye movement information.

All in all, the present results provide evidence that the entorhinal cortex represents conceptual affordances. These findings should lend weight to a link between structure learning and behaviour, in particular gaze behaviour. In his seminal work on visual perception, James Gibson eschewed the idea of cognitive maps because of the lack of an "internal perceiver to read [the] map". (Gibson, 1979) Our work suggests that, on the contrary, cognitive maps could provide a normative guide to navigation through action representation. In the next chapter, we will seek to extend this finding by exploring how action representations apply across different relational structures.

Cross-context rotation of visuospatial action representations in the human entorhinal cortex

3.1 Abstract

Switching between action plans is crucial to adapt to changes in our environment. In the previous chapter, we showed that the entorhinal cortex, a key neural substrate for cognitive maps, is sensitive to learned action possibilities. Nonetheless, it is unclear whether these action representations generalise across different contexts or are tied to a specific map. In this study, we teach participants to navigate around a visual map using eye movements in two different task contexts. Across all states and within contexts, entorhinal cortex patterns are more similar when possible actions are more aligned. Across contexts, action similarities are highest when action directions are rotated in one graph context prior to computing action alignment. This suggests that human representations of relational structure contain a representation of action which rotates uniformly across different tasks, allowing both differentiation of task contexts through rotation and a generalisation of distance information via a shared representational format.

3.2 Introduction

Similar environments may afford radically different actions in different times or contexts. To sail the same course every day, a sailor must adjust the sails and rudder based on the wind and currents. Making a coffee in two different kitchens may entail a completely different set of actions, albeit with a similar result. Nonetheless, it is unclear how action information is integrated into structural knowledge about the world across different task contexts.

The hippocampal formation contains a variety of specialised cell types which appear to respond selectively to features of space, such as location, orientation or proximity to a border or object (Barry et al., 2006; Solstad et al., 2008; Deshmukh & Knierim, 2013; Høydal et al., 2019; Bicanski & Burgess, 2020). In recent years, research has demonstrated that these neural populations are also responsive to non-spatial dimensions, such as pitch or value (Aronov et al., 2017; Knudsen & Wallis, 2021; Qasim et al., 2023). In primates, visual coding appears to be particularly prevalent: in monkeys, entorhinal cells map out positions in visual space with the same hexagonal grid characteristic of spatial grid cells (Killian et al., 2012, 2015; Killian & Buffalo, 2018); in humans, similar hexadirectional responses have been measured using non-invasive neuroimaging techniques in both visuospatial and non-spatial domains (Constantinescu et al., 2016; Viganò & Piazza, 2020, 2021; Knudsen & Wallis, 2021; Park et al., 2021; Viganò et al., 2023; Nitsch et al., 2024; Sigismondi et al., 2024; Barnaveli et al., 2025).

Different regions of the hippocampal formation appear to show a functional distinction in how they respond to new environments. Place cells in the hippocampus, which typically fire at one or few specific locations in space, often remap to fire in new locations when the environment is changed (O'Keefe & Burgess, 1996). On the other hand, cell populations in the entorhinal cortex tend to maintain the same relationships: although the orientation of grid or vector cells often rotate uniformly in new environments, they typically maintain the same cellwise relationships and fire at the same periodicity (Hafting et al., 2005; McNaughton et al., 2006; Fyhn et al., 2007). This observation has led to suggestions that entorhinal representations are used in aid of generalisation, as the same metric relationships can be used across different environments (Behrens et al., 2018; Whittington et al., 2020). In support of this, human neuroimaging experiments have shown that fMRI responses in the entorhinal cortex are more similar when tasks have the same underlying structure (Baram et al., 2021; Mark et al., 2024).

Nonetheless, there are indications that entorhinal cells are responsive to specific environmental changes, beyond rotation. In rodents, it has been shown that grid cells distort around goals, either increasing firing or clustering around goals (Boccara et al., 2019; Butler et al., 2019; Viganò et al., 2023). In hairpin mazes, not only do rodent grid cells splinter to represent specific arms of the maze, but they remap when the maze is moved to a new room (Derdikman et al., 2009; Whitlock et al., 2012). Moreover, periodic cells in the entorhinal cortex align to landmarks, both visual landmarks in a VR maze (Wen et al., 2024) and to imagined stimulus positions in a mental navigation task (Neupane et al., 2024). In humans, hexadirectional responses in fMRI are disrupted in special populations with different lived experiences, such as blind people (Sigismondi et al., 2024). One interpretation of these results is that entorhinal representations are sensitive to the action distribution of an environment; changing the afforded actions changes how neural populations map out space (Jeffery, 2021; Ginosar et al., 2023). Indeed, in the previous experiment we showed using a functional neuroimaging paradigm that neural patterns in the right entorhinal cortex were more similar when afforded actions were more similar. While previous results in rodents have largely centred around grid cells, it is possible that related cell populations such as object vector cells and, more generally, action representations, may also distort in a context-specific manner (Whittington et al., 2022). If maps contain an 'entangled' representation of action, this may be beneficial behaviourally: simply retrieving a map provides the appropriate, context-specific actions that may be taken.

In order to generalise, however, it may be important to separate action from the specific task context. Computational work has proposed that action representations may form a separate 'factorisable' substrate (Bakermans et al., 2025), independent of representations of task structure. Models of the hippocampal formation often treat action as a separate input which transforms between states (Stachenfeld et al., 2017; Whittington et al., 2020; Sorscher et al., 2023) or performs a vector movement on a grid cell manifold (Chandra et al., 2025). This is ideal for generalisation: in a new environment, an agent has to simply implement already-known actions and can re-use the same neural machinery. In comparison to an entangled representation, it may take longer to learn and not fully explain how actions relate to the distortions of the grid manifold mentioned above. In general, this presents a dichotomy between two types of action representation in the entorhinal cortex: entangled or factorised (Whittington et al., 2022).

In this study, we first aimed to replicate previous results showing action representations in the entorhinal cortex, and then tested these two possibilities using a visual exploration paradigm in which participants learned to associate animal images with certain eye movements in two different graph conditions. The neural representation of each animal across all states, then within and across each context was then tested using functional magnetic resonance imaging (fMRI). Within the same graph context and when considering

all states, neural representations in the right entorhinal cortex were more similar when action possibilities were more closely aligned, replicating our previous finding that the action information is incorporated into entorhinal representations of task structure. Across different graph contexts, neural action representations were most similar when actions were rotated uniformly across subjects, in line with prior studies demonstrating a uniform rotation of entorhinal representations in new environments or contexts.

3.3 Materials and methods

As discussed in the introduction, the present study had two key objectives: firstly, to replicate the finding from the previous study that the entorhinal cortex represents action information; secondly, to explore whether action representations are linked to a specific graph context or generalise across different graph structures.

As before, we needed to way to compare actions across different stimuli. Whereas in the previous experiment we were aiming to probe actions in a conceptual domain, for this experiment our hypothesis simply required a uniformity of action (concrete or otherwise) across different states. Given that the previous experiment had shown an embodiment of action in eye movement, we chose to directly operationalise action as eye movement, such that the same movement on a screen could be considered as the same action. This was independent of screen position or visuo-semantic features of the task: an eye movement right could be considered similar (in screen coordinates) regardless of where it started from. Therefore, we would be able to test state pairs in terms of whether their afforded eye movement overlapped, with the prediction that more similar afforded eye movements would correspond to more similar neural patterns in the entorhinal cortex. An additional advantage of this choice was that we would be able to accurately distinguish representations of planned (or imagined) action from enacted action, which is not possible in a conceptual space. This will be discussed further later in this chapter.

Following the logic of the previous experiment, we then needed to teach state-eye movement associations in two different relational structures and then probe their representation using fMRI. This presented a challenge: how to teach an eye movement? After some initial piloting, we settled on a ‘spotlight’ style design, in which stimuli were only revealed when the participant looked in the correct screen position. To ensure learning of the correct eye movement, animal images (states) would only be revealed when the participant looked between adjacent images according to an underlying graph structure. By changing the graph structure between blocks but keeping the images and positions consistent, we could then teach two different sets of eye movements for each image. This allowed us to test both questions: firstly, if entorhinal neural pattern similarities scaled with action similarities, and secondly, whether entorhinal representations of action generalised across graphs or not.

Therefore, we aimed to teach participants how to navigate between masked animal images and then probe the representations of these images using fMRI. In the following section, I will describe the task design and

experimental setup before moving on to data collection and a run-down of analysis methods used. Many of these are shared between experiments, but I will nonetheless re-state analysis details for clarity.

3.3.1 Task

Experimental design

The experiment consisted of an eyetracking task split into two consecutive days. Participants were informed that they had to play the role of a new hire at a zoo, who was tasked with feeding the animals. This feeding order varied according to the shift (which was designated as 'day' or 'night'). This fictitious cover story ensured that participants understood that although they had to learn two separate transition sequences, the visual and semantic features of the individual elements remained constant.

Both days involved two sessions, in which participants first learned the layout of the zoo by day and by night before being tested on their knowledge of adjacent animal transitions. In the training ('visual exploration task'), participants learned how animals were connected by freely looking over the screen until adjacent animals in the underlying graph were revealed. In the test ('saccade task'), participants were instructed to observe individual animals before carrying out the relative eye movement to get to a subsequently-presented target animal. This allowed us to measure neural responses to individual animals, as well as the enacted action between start and target animal.

The first day's experimental session lasted two hours, of which the training lasted approximately an hour (8 blocks) and was followed by twenty minutes (2 runs) of the saccade task. On the second day, a shorter training session of around half an hour (4 blocks) preceded the scanner session, during which participants carried out 9 runs of the saccade task. The first run was carried out during the anatomical scan as a practice, while the subsequent eight runs were performed during functional runs. In total, the scanner session typically lasted an hour and a half, including eyetracker set-up and calibration. At the end of the experiment, participants were asked to fill in a brief questionnaire in which they had to describe their strategies, sketch the zoo by day and night and fill in relevant demographic information (age, gender, education level).

Stimuli

The stimuli were designed to be engaging and easily recognisable, given the fast-paced design and lower quality image presentation in the fMRI scanner. Therefore, animal images on a coloured background were

used. Animal images were drawn from a dataset from Possidónio and colleagues (Possidónio et al., 2019), and particularly arousing or visually confusing stimuli were manually excluded by the experimenters (e.g. spiders, tropical fish). Of the final 39 animal images, 12 were randomly selected for each participant and randomly associated with one of 12 colours and 12 positions. The white background from the original images was removed and replaced with a circle of the randomly-assigned colour for each participant. This randomisation ensured no effect of pre-existing associations or visual effects at the group level, while creating recognisable images and leaving space for future exploratory within-participant analyses based on semantic similarities between animals or colours. Images were displayed at a width of 80 pixels.

Graph structure

The underlying graphs described how participants were able to transition between different states, with the aim that participants would learn both task context as part of a looped graph and the affordances associated with individual images. The two graphs used (Figure 1, left) were designed based on a hexagonal grid, allowing six possible actions from each state. Each state afforded two of six possible actions within each graph, creating a looped structure. Across the graph, each action was repeated four times. As mentioned above, stimuli were randomly assigned to a state on each graph for each participant.

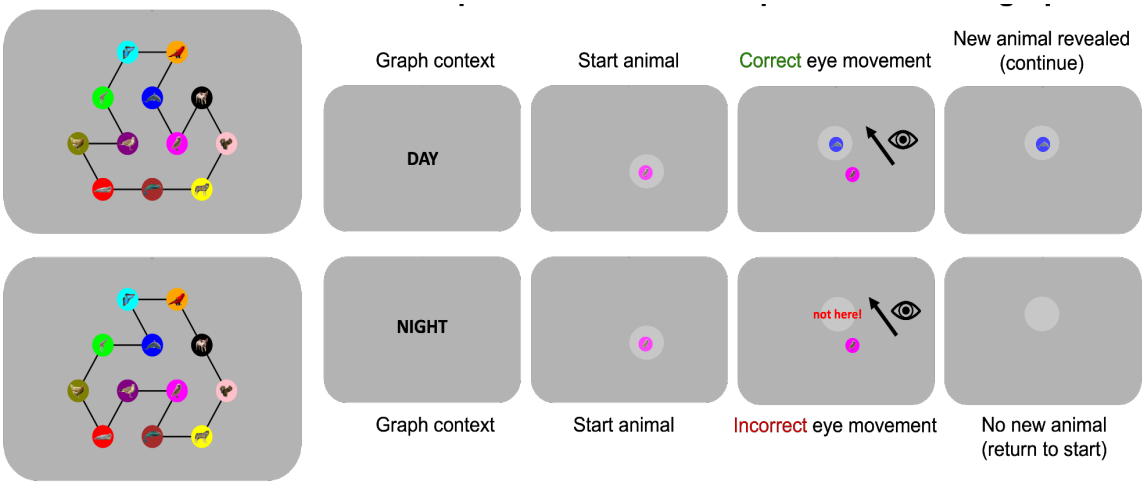

Figure 1: Participants learned to navigate two graphs using an ocular 'spotlight'. *Left: graph structures for each context. Right: Trial timeline. participants were presented with a random image from the graph in isolation and then explored the screen using eye movements. Animals were revealed when they looked along from the current animal to a possible successor according to the graph structure.*

Visual exploration task

The aim of the visual exploration task (Figure 1, right) was to teach participants the layout of each graph using a 'spotlight' approach whereby animal images were only revealed when the participant looked at the correct screen position, in the correct order. Training consisted of a blocked design, in which participants learned the structure first for one graph and then the other. During the first day, participants carried out 8 blocks of training (4 of each graph), and an additional 4 blocks during the second day (2 of each graph). The order was balanced across participants, to ensure no group-level primacy effects.

Before the task started, eye tracker calibration and validation were run. Calibration was repeated when participants looked away from the chin or forehead rest; otherwise at the start of each block a drift correction was performed. Before the main task started, participants were presented with white dots on the screen marking the animal positions and chose with a button press when to start the main experiment. Each block consisted of 12 trials.

As mentioned above, learning followed a 'spotlight' paradigm whereby images or cues were only presented when the participants looked at the relevant position on the screen. At the start of each trial, central white text indicating the shift ('day' or 'night') was followed by the presentation of an animal image in the correct position according to the predefined assignment for that participant (see 'Stimuli'). The rest of the screen was blank (grey), hiding the other animals. After fixating the initial image, participants then moved their eyes until either another animal image appeared (indicating a correct transition within that graph) or writing ("not here") indicated they had looked at the position of an incorrect animal for that graph. Animals disappeared after 1s without being fixated.

After participants had found all 12 animals of the graph (corresponding to a complete loop) and returned to the start animal, a new trial started. Trials were paired, such that the next trial started with the same animal, and the participants were encouraged to follow the graph in the opposite direction the second time, in order to learn both possible actions. All possible start animals were presented an equal number of times for each graph across training, both on day 1 (96 total trials; 4 of each animal per graph) and day 2 (48 total trials; 4 of each animal per graph).

The trials were self-paced. In order to assess where participants were looking, we sampled their gaze position from the eye-tracking system at a frequency of 1000Hz and compared it to the position of the animals on the screen. If the fixated point was within 60 pixels of the centre of an image, it was considered

a selection of that animal. As this covers a region outside the image itself, this ensured both that any minor inaccuracies due to the eye tracking system (e.g. drift) could not make the task impossible, and that it was possible to find the animals in early trials without good knowledge of specific animal positions.

Saccade task

The aim of the saccade task (Figure 2), as with the recall task in the previous experiment, was to isolate neural and ocular responses to individual animal stimuli. These could then be compared in isolation according to action similarities both within and across graphs, in line with our experimental questions. Whereas in the previous experiment participants were presented with a continuous sequence of images interrupted by occasional probe trials, we reasoned that the increased difficulty of the task meant participants may easily lose focus during the scanner session. Therefore, we introduced a saccade-to-dot task after every animal presentation, whereby participants were asked to move in the relative direction of a target animal compared to the presented animal, following the learned graph. If the transition was impossible, they had to keep fixation on the target animal.

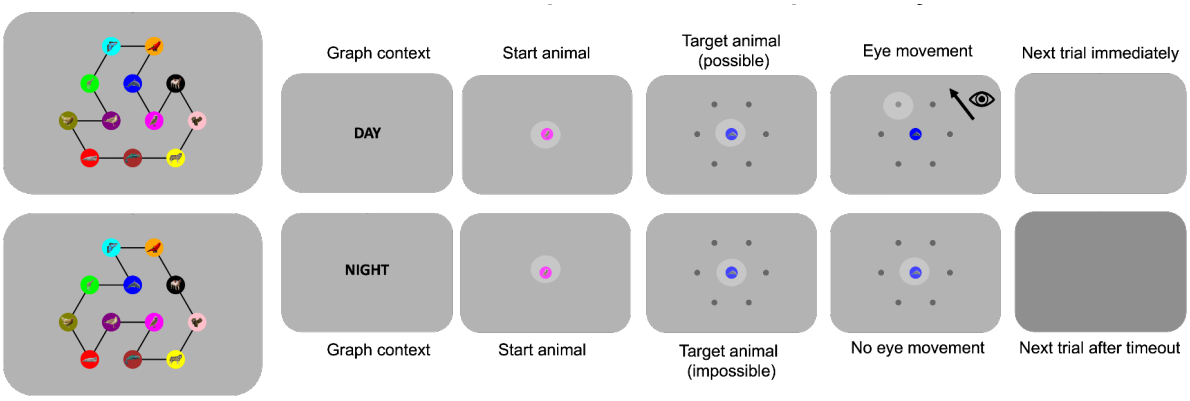

Figure 2: Participants selected the correct action from each animal image according to graph context. *Left: Graph structures for each context. Right: Trial timeline. Participants were asked to view a 'start' animal image in isolation, and imagine possible successors. They were then presented with a 'target' animal, upon which they had to either (1) select a dot in the relative direction of this animal compared to the 'start' animal using an eye movement (task trial), or (2) if the target was not possible within that graph, keep fixation on the centre (decoy trial).*

On the first day, participants were presented with two runs of the saccade task, and on the second day nine runs, of which eight were carried out while functional MRI data was acquired while the first was carried out during an anatomical scan before the functional session started. Each run took approximately eight minutes. Within each run, trials were divided into four interleaved blocks of 'day' or 'night'. This was indicated by white text with the name of the condition displayed for ten seconds before each block.

Each trial started with a fixation cross (mean 2.5s, jittered by drawing from a truncated exponential distribution), followed by a 'start' animal image (2s). After another inter-stimulus interval (mean 2s, jittered by drawing from a truncated exponential distribution), a 'target' image was presented (1.2s) surrounded by 6 dots which corresponded to possible action directions. Participants had to make an eye movement (typically but not necessarily a saccade) from the position of the central animal to the dot which represented the relative position of the target. Each run contained 12 decoy trials, in which the target image was drawn from the wrong graph. In these cases, participants were asked to maintain fixation on the target animal. After a dot selection, the script showed a blank grey screen for 0.3s before starting the next trial. In the case of timeout, the screen changed to a darker grey for 0.3s before continuing.

The actions were balanced across the run, such that each run consisted of eight of each possible direction. The central position was slightly jittered spatially by selecting randomly from one of four points corresponding to points on a square around the centre (balanced by condition). This ensured that any neural effects were caused by the relative action, rather than the screen position of the target dot. Moreover, the graph was designed such that affordance-based state similarities were not correlated with a matrix of eye movement response similarities (Spearman rho < 0.1), thus ensuring that any resultant effects could not be driven by oculomotor movement.

To avoid any temporal autocorrelations in the data, the order of trials, ITI distribution and assignment of position and action were permuted until a combination was found for which the distance between adjacent trials did not strongly correlate (Spearman correlation < 0.1) with any model-based distances (see below; RSA) or screen distance between jittered fixations. The block order was switched between runs, such that alternating runs would be DAY-NIGHT-DAY-NIGHT or NIGHT-DAY-NIGHT-DAY, in order to ensure no systematic changes in signal between blocks due to temporal drift.

Each run consisted of 60 trials, of which 30 were drawn from each graph. Although participants learned twelve animal images, only a subset of six animal images were used for final analyses to increase signal-to-noise by allowing more repetitions of each condition. To ensure that participants nonetheless maintained

a good memory of the entire graph structure, both subsetted trials (48; 4 repetitions of each condition) and non-subsetted trials (12; 1 repetition of each condition) were presented. On the first day and the first run in the scanner (during the anatomical scan), this was reversed so that 48 images were drawn from the non-subsetted group and 12 from the subsetted group. This avoided overtraining on the images used in the scanner.

3.3.2 Data

In the following section, I will describe data acquisition procedures. These are broadly similar to the previous experiment, with the exception that eyetracking data was collected both during behavioural training and in the MRI. On the first day, participants learned the task through visual exploration, and then carried out two practice runs of the saccade task. On the day after, they carried out four blocks of the training task before repeating the saccade task in the MRI scanner.

Participants

Fifty-three healthy German-speaking volunteers participated in the study. All participants were right-handed, with normal or normal-to-corrected vision and reported no history of neurological or psychiatric disorders. All participants provided written informed consent to participate in the study, and were compensated for participation. The study was approved by the local ethics committee at the University of Leipzig, Germany (protocol number 400/22-ek).

Two participants were excluded due to a mean accuracy of over one degree of visual angle across scanner runs, meaning that they were unlikely to have been able to engage meaningfully with the task due to eyetracker inaccuracy. A further two participants were excluded due to performance lower than two standard deviations from the mean. These were excluded from all analyses, resulting in a final sample size of forty-nine. 27 self-identified as female, with an overall mean age of 26.2 (s.d. 4.0).

A target sample of over 50 participants was set during data collection, to match the previous study (which was sufficiently powered to find action representations in the entorhinal cortex). As before, we also expected many exclusions due to poor performance or eyetracking quality, but found that most participants

were able to perform the task well (although with a notable increase in variability relative to the previous experiment).

Eyetracker Data Acquisition

For eyetracking data gathered outside of the MRI scanner, participants were seated in front of a monitor leaning on a chin rest and forehead bar, at a distance of 57cm from the infrared camera and receiver. We recorded monocular gaze position continuously using an EyeLink 1000 Plus (SR Research), at a sampling rate of 1,000 Hz.

The same system was used to record eye data during the MRI scanner session, but in this case the participants were lying down and an eyetracker-compatible mirror positioned on the head coil was used to reflect the infrared signal at the appropriate angle. In this case, distance from the camera varied as a function of head size.

To ensure accurate recording, we calibrated the system at the start of every session using the built-in calibration and validation protocols from the EyeLink software with nine fixation points, and re-aligned using a central fixation point at the start of each run. Fixations were accepted if they were less than one degree of visual angle from each point. Accuracy (mean distance from fixation points) and reliability (standard deviation in distance from points) on each day were calculated using the distance from the central fixation during realignment (day one accuracy: mean 0.46 degrees, standard deviation 0.20; day one reliability: mean 0.2 standard deviation 0.17; day two accuracy: mean 0.5, standard deviation 0.15; day two reliability: mean 0.28, standard deviation 0.12).

MRI Data Acquisition

Given that our previous study was able to detect action representations in the entorhinal cortex, we used a similar procedure for MRI data acquisition in this experiment. MRI data were recorded using a 3 Tesla Siemens Magnetom Skyra scanner (Siemens, Erlangen, Germany) with a 32-channel head coil. The scanning session started with an anatomical scan, which used a T1-weighted MPRAGE (TR = 5000ms, TE = 19.6ms, voxel size = 1 mm isotropic).

Following this, blood-oxygen-level-dependent (BOLD) contrast was measured for eight runs of the saccade task using T2*-weighted whole-brain gradient-echo planar imaging (GE-EPI) with a multiband acceleration factor of 5. As with our previous work (Eperon et al., 2025) and suggestions from other studies (Nau, 2022), the acquisition box was aligned with the long axis of the hippocampus in order to optimise signal-to-noise ratio in the hippocampal formation.

The fMRI sequence had the following parameters: TR = 1000ms; TE = 22ms; voxel size = 2.5mm isotropic; field of view = 204mm; flip angle = 62°; partial fourier = 0.75; bandwidth = 1794 Hz/Px; multi-band acceleration factor = 5; 65 slices interleaved; distance factor = 10%; phase-encoding direction: A -> P. Each run lasted a different length due to variation in response times, and was thus stopped manually when the task finished (around 8:20 minutes).

At the end of every run, fieldmaps were acquired to measure and correct for inhomogeneities in the magnetic field. Fieldmaps were acquired using opposite phase-encoded EPIs, using the same voxel sizes and field of view as the functional runs with TR = 8000ms; TE = 50ms; flip angle = 90°; partial fourier = 0.75; bandwidth = 17.934 Hz/Px.

Stimuli were projected into a screen situated behind the participant, which was viewed using an eyetracker-compatible mirror mounted on the head coil. Ocular responses were measured using a scanner-compatible eyetracker with an infrared camera (Eyelink 1000 plus, SR Research), which was positioned below the screen (see above for details).

Preprocessing of fMRI data

Preprocessing was carried out using fMRIprep 22.0.1 (see Appendix for a boilerplate provided by fmriprep, which is the same for both experiments). The processing steps were the same as the previous experiment, and included segmentation, head motion estimation and correction, slice timing correction, co-registration to anatomical images, and fieldmap-based (AP-PA) susceptibility distortion correction. All analyses were carried out in volumetric space, using subject space (T1w) for all analyses. Before running general linear models (GLMs), volumetric data was smoothed using a Gaussian FWHM filter with a 5mm width.

3.3.3 Analyses

As described in the introduction, the present experiment had two goals. Firstly, we aimed to replicate the result from the previous experiment in which we found that neural pattern similarities in the entorhinal cortex scaled with action similarity. Secondly, we aimed to test whether action representations generalised across contexts or not. The analyses carried out follow the same broad pattern as the previous experiment, first aiming to establish that participants could do the task before isolating neural representations in the entorhinal cortex using ROI-based RSA and testing them against models of state similarity based on action similarity. As before, we ran RSA in each hemisphere of the entorhinal cortex.

Nonetheless, there are some key differences which will be explained in more detail below. As we aimed to compare within and across contexts, we separated neural and model RDMs into two halves which compared states only within the same graph, or across graphs. An exploratory analysis also involved testing for a rotation of graph structure across context; this methodology will be described in the main results section.

Finally, it should be noted that the results presented in this chapter are only the first part of a longer investigation which will aim to understand whole-brain effects and representation of enacted oculomotor action in the medial temporal lobes. One particularly evident omission is analysis of eyetracking data, which has been used as a performance measure without any comparisons to neural data. Where pieces are missing, I will highlight the future analyses which will be done to better understand our data in the results and discussion sections.

Analysis of behavioural data

Firstly, we aimed to understand if participants had indeed learned the two graphs. Performance in the exploration and saccade tasks was measured using, in both cases, proximity of eye position to targets during the task. In order to do this, eye position was continually measured during the task. When an eye position within 80 pixels of the centre of an animal image or target dot was registered (outside of the scanner, and the first ten scanner subjects; inside the scanner for most subjects a radius of 60 pixels was used to reflect the smaller screen size), either the feedback of an animal image or text was revealed (exploration task), or the trial ended (saccade task).

For a single trial, performance in the exploration task can be understood as a percentage of movements within one loop of the entire graph which alighted on the correct animal. In the saccade task, performance

within each trial is a binary measure indicating if the correct dot was fixated, or lack no dot fixated in case of decoys.

As mentioned, a correct response in the exploration task indicated that participants moved their eyes to an 'adjacent' animal in the graph, rather than another animal (for example, a transition from the incorrect graph, or a longer movement to a distant animal). Only responses which alighted on a possible animal image were considered. This meant that the chance threshold varied based on spatial position and position in the trial, as some positions have more adjacent animals than others; therefore, a chance level of 0.702 for any trial was calculated by taking the mean performance across all transitions and all subjects within the first trial of the experiment.

To check learning, performance on the second day was compared both to chance and to performance on the first half of the first day, in both cases using the average performance across all transitions and trials for each subject. A Wilcoxon signed-rank test was used with an alpha level of 0.05.

For the saccade task, performance was taken as the percentage of responses which either alighted on the correct dot around the target, or successfully stayed in the centre (in the case of decoy trials), for a chance level of 1/7. Again, performance against chance was measured using a Wilcoxon signed rank test with an alpha level of 0.05. As mentioned above, participants with average performance lower than two standard deviations from the mean during the scanner task were considered outliers and were excluded from all analyses on both days. The first run during the scanner (while anatomical data was captured) was excluded from all analyses, as participants were explicitly informed this was a practice run. As eyetracking data in the scanner was notably noisy, no trials were excluded from the neural analyses based on performance.

To be sure that participants were using graph knowledge instead of spatial position, we checked that they were more likely to make no response in the case of decoy trials (targets from the wrong graph) relative to trials where a response was necessary. This was calculated using d-prime (where no response in the case of a decoy was considered a hit, and no response in the case of other trials was considered a false alarm), and the resulting d-prime values at group level were compared to 0 using a Wilcoxon signed rank test with an alpha level of 0.05.

Finally, to ensure that there were no differences in reaction time or performance as a function of state identity during the scanner session, a repeated-measures ANOVA was run, using image identity, graph identity and the interaction of the two. An alpha level of 0.05 was used.

Representational similarity analysis (RSA)

As before, the aim of RSA was to isolate the neural signals found in each hemisphere of the entorhinal cortex, and compare these to model RDMs based on action similarity.

Following preprocessing, functional data was smoothed using a 5mm FWHM kernel and beta maps for each condition of interest were extracted by running a GLM which contained state-graph combinations as conditions (modelled using a boxcar function). This resulted in 24 conditions, although only 12 were used in the RSA analyses due to the subsetting approach mentioned above. Other regressors included six motion regressors from fmriprep, an intercept, and regressors corresponding to other visual or motor features of the task.

In order to compare neural pattern similarities within the entorhinal cortex, first we extracted voxel patterns for the region (in each hemisphere) using subject-specific ROIs. These were taken from maximum probability maps from Juelich Atlas 3.0.3 and mapped to subject space using ANTs 6.0.1 using the MNI-to-subject space transformation generated during fmriprep preprocessing. in an additional control analysis, we used two larger bilateral and right-lateralised MTL regions-of-interest, which included the hippocampus, parahippocampal gyrus, and entorhinal cortex.

As before, voxel patterns within each region-of-interest were then compared across conditions to create neural dissimilarity matrices using cross-validated Mahalanobis (crossnobis) distance, using residuals from the GLM to create an estimate of noise. This produced 12x12 neural RDMs for each subject and region. As will be discussed below, these were then compared to model-based predictions of neural pattern similarity based on action similarity.

Representational similarity analysis: model matrices

To test for action similarity, we constructed model RDMs based on the similarity between actions afforded by pairs of images. We reasoned that there may be two possible formats for action representation. Firstly, actions may be compared by identity: a movement east may be considered equally different from a movement west or a movement north-east. Alternatively, the angle of the action may be important: a movement east might be more similar to an action north-east than an action west. I refer to these two models

as ‘action identity’ or ‘action distance’, and their operationalisation will be described in detail below and visually in Figure 3. It is worth noting that in the previous experiment, both models would have produced the same model RDMs.

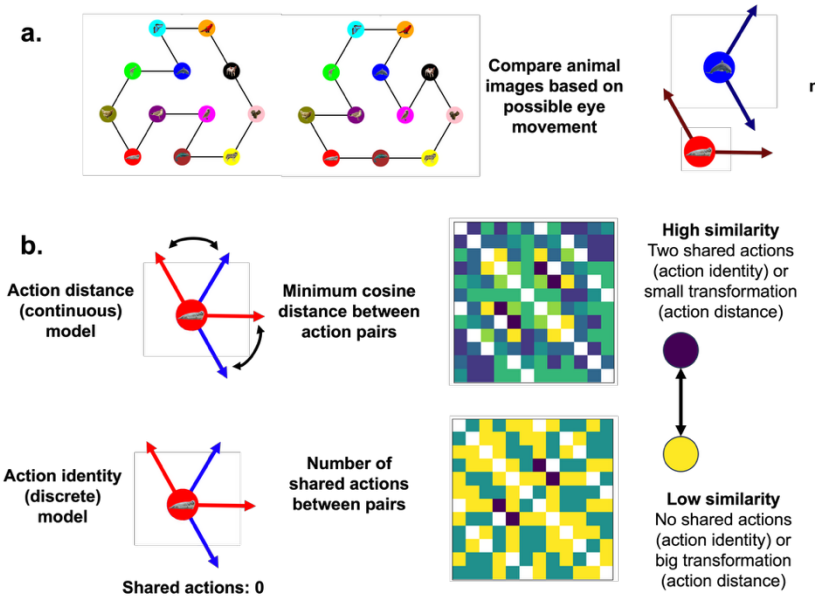

Figure 3*: Models of state similarity. a. To construct model RDMs, we compared animal images based on the possible eye movements they afforded. b. We considered two alternate models: a model based on action distance, modelling the angle between action pairs for the two states (above), or a count-based model based on the number of shared actions across state pairs (below).*

To compare action identity, I calculated count-based RDMs based on the number of shared actions between pairs of states, subtracted from two (the maximum number of shared actions). Therefore, a high value indicated no shared actions (predicting a high neural pattern distance), whereas a value of zero indicated two shared actions and therefore a low neural pattern distance. It should be noted that most states shared only one or no actions due to the layout of the graph, with only one pair of states (across graphs) sharing the exact same actions.

As mentioned above, the action distance RDM aimed to capture the dissimilarity between action pairs in terms of how dissimilar the action angles were. This presented a logistical problem. To map between two angles, cosine distance can be used. However, each state afforded two actions, meaning that we needed to compare two pairs of actions (for the sake of simplicity, let us imagine we are comparing two pairs of

actions labelled A,B and C,D). This provided two possible mappings between action pairs (e.g. A-C and B-D or A-D and B-C), either of which offered a possible way to compare the angle-based similarity of action pairs.

To solve this problem, we assumed an efficient mapping between state pairs whereby the mapping which produced the lower average distance was taken. This intuitively reflects how we might expect similarities to be calculated. This may be illustrated with a simple example. Suppose action pairs A,B and C,D describe actions going left and right, respectively. Intuitively, we should expect these pairs to be considered very similar, as they describe the same actions. However, this is only the case in one mapping between actions (A-C and B-D), while the alternative mapping (A-D and B-C) produces a high cosine distance of 2 (corresponding to 2(1-cos(180))/2). The same is the case for a variety of other examples, such as action pairs where only one action is aligned or both differ by a small angle. Alternative approaches were considered during the design phase (e.g. choosing the mean angle of each action pair or averaging the distance from the two potential mappings), but the minimal distance approach was considered the simplest and most intuitive and so only this was tested.

Therefore, both possible mappings were considered (e.g. A-C and B-D or A-D and B-C), and for each mapping the average cosine distance between action pairs was taken. The mapping with the lowest average cosine distance was used for the distance between states. This produced a continuous RDM (action distance RDM), which captured action similarity in terms of the different angles of the actions afforded by different states.

To carry out within- or cross-graph comparisons, the resulting representational dissimilarity matrices (RDMs) were subsetted so that they only included either squares where both conditions were drawn from the same graph (within) or different graphs (across), as shown in Figure 3.

A possible confound may have derived from cross-graph comparisons between animal images in the same position during learning. As images in the same position shared the same visuo-semantic qualities, this may artificially drive high similarities. Therefore, these squares of the RDMs were excluded from all analyses, such that we never compared neural or model distances between the same animals. This can be seen in the model RDMs, which show ‘empty’ white squares in both the diagonal and for 6 squares in the cross-graph. Both across- and within- graph matrices consisted of 30 squares, while the full RDMs consisted of 60 squares.

An additional confound may have arisen from the trial-wise responses, which could have elicited the same eye movement across trials with similar action possibilities. This was controlled for during the experimental design by creating two model RDMs using state-action combinations as conditions. One RDM predicted condition similarities in terms of response (enacted action) similarity, while the other predicted similarities in terms of the actions afforded by each state (as for the action identity matrix above). The two were not correlated (Spearman rho < 0.1), indicating that our effects were unlikely to be driven by enacted action similarity during the response phase.

Finally, control matrices were constructed based on the Euclidean distance between states during learning (continuous; both within and across graphs), or the minimum link distance in the graph structure (count-based; within graph only). These were used in a partial correlation (see below) to ensure that any resultant effects could not be explained by other features of the learned graphs. These controls accounted for possible effects due to the position of the animal during learning in terms of screen distance (Euclidean distance), as well as possible effects related to graph distance or transition probabilities between states (link distance).

Representational similarity analysis: statistical tests

For statistical tests, model matrices were compared to neural dissimilarity matrices using tie-corrected spearman correlation.

As several model matrices were correlated with each other (notably, the action identity and action distance models were correlated at 0.51 and the action identity and link distance models at 0.32; Spearman's rho), a partial correlation was used to exclude the effect of other models. In all cases, the partial correlation was calculated using a recursive formula based on rank correlations. In cross-graph comparisons, the Euclidean distance matrix was controlled for, while in both within-graph and all state comparisons, both Euclidean distance and link distance were excluded. For the whole-graph comparison using action distance and action identity models, each model was also excluded from the other by including them in the partial correlation.

The resulting correlation values were approximated to a normal distribution using a Fisher Z transformation, before group-level tests were carried out using one sample, one-tailed t-tests (as standard for the implementation in RSA toolbox (Nili et al., 2014); see also the considerations in the previous chapter). Multiple comparison correction was carried out using Bonferroni correction, resulting in alpha thresholds of 0.0125 for whole-matrix comparisons (two hemispheres, two models each), and 0.025 for the subsequent

within and across comparison in the right entorhinal Cortex (two models). Comparison between the within and across models was carried out using a two-tailed t-test, with an alpha level of 0.05.

3.4 Results

In the following section, I will describe the findings of the analyses described in the previous section. First, I checked that participants were able to learn the two graph structures, meaning that they were able to perform the correct eye movements in each graph context. Having established that participants had learned the task correctly, we tested whether the entorhinal cortex represented action possibilities using a targeted ROI-based RSA approach, thereby replicating our findings from the previous experiment. Following this, we tried to distinguish between a context-specific or context-general representation of action by comparing action representations within or across graphs. As the results were somewhat inconclusive, I tried an exploratory analysis based on rotation of action orientation across contexts, drawing on previous work showing cross-environment rotation of entorhinal representations. As with the previous chapter, all results are presented with raw p-values, and significance testing follows the Bonferroni corrected alpha values reported in the previous section.

3.4.1 Learning of ocular affordances in different graphs

We predicted that participants would be able to learn state-action associations in two different graph contexts, despite each state sharing the same visual features and position across graphs.

Participants performed a visuospatial navigation task in which animal pictures were arranged on certain positions on the screen, but only visually presented when directly fixated (‘visual searchlight’) in the correct order according to an underlying graph. This setup was used to teach them to conduct specific eye-movements (actions) for each picture (state) to make the next picture appear. To teach the two graph contexts, they were informed that they had been hired at a zoo and had to learn the feeding order of animals in two shifts, day and night (Figure 4a left).

In practice, participants were first cued with a given graph context (‘day’ or ‘night’). Within each graph block, they were first presented with an individual animal on an otherwise-grey screen, and asked to look around until they saw another animal. Animals which were considered correct according to the graph were revealed when participants looked in their direction, while incorrect positions showed the text “not here” and participants were instructed to return to the last previously-seen animal (Figure 4a right).

To give a concrete example of the task, each trial started with cueing of shift information (day or night). Then, a participant might see a flamingo presented on the centre-right of the screen, and no other visible animals. The participant looks at the flamingo to start the trial. If the participant then moves their focus away from the flamingo, the flamingo will disappear. If the participant looks up-left towards the location of the hidden dolphin image, one of two things may happen. If this is a correct animal according to the graph structure, the image will appear. If this is not a possible transition, text will appear saying 'not here!', thereby indicating that the participant should look back toward the (now-invisible) flamingo to reveal it again and look for another transition.

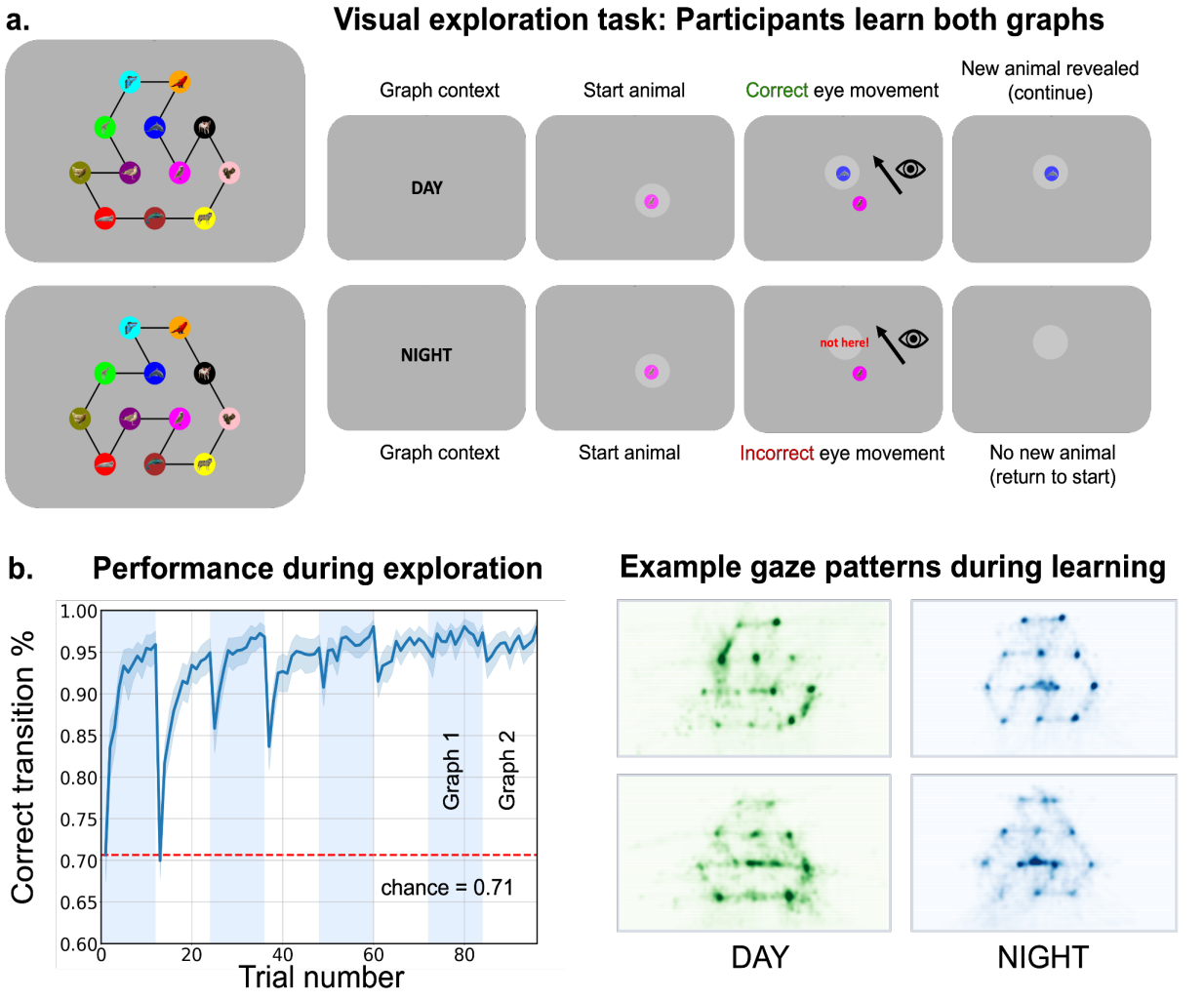

Figure 4: Participants learned to navigate two graphs using an ocular 'spotlight' *. 1a. Trial timeline: participants were presented with a random image from the graph in isolation, and then explored the screen using eye movements. Animals were revealed when they looked along from the current animal to a possible*

successor according to the graph structure (4a left). 4b left: Participants learned the graphs by the end of the first day's training through interleaved blocks from each graph context. 4b right: example heatmaps from the first run for each graph for two example subjects.

If participants were able to learn the graph, at each step through the zoo they should have been able to look in the direction of adjacent 'correct' animals rather than 'incorrect' animals within each block. We therefore compared the percentage of correct transitions within each trial in the final two blocks to a baseline using the percentage of correct responses in the first trial of the first blocks across all participants (0.706, reflecting differential chance levels based on image position and already-seen images within one trial). As expected, participants were able to learn the graph structure (wilcoxon(49)=1225, $p<0.001$). This was not driven by learning in the first trial of each block: performance on trials where blocks switched was greater for the final two blocks than the first two blocks (wilcoxon(49)=1160, $p<0.001$). The time to select a new image decreased correspondingly, from a mean of 1.58s in the first two blocks to 1.20s in the last two blocks (wilcoxon(49)=144, $p<0.001$), and again this decrease was seen also in switch trials for those blocks (wilcoxon(49)=125, $p<0.001$). Therefore, participants learned the graph structure well, and were able to switch easily between contexts.

3.4.2 Learning of state and graph-specific eye movements

We then tested whether participants were able to enact eye movements based on state and graph information, independently of positional information. To do this, participants were first presented with a 'start' animal and then asked to make a saccade in the direction of a successively-presented 'target' animal, which corresponded to the position of one of six dots around the target. If the target animal was not a possible transition ('decoy'), participants were asked to keep fixating the target (Figure 5a).

If participants had indeed learned the graph-state-action associations correctly, we predicted they would be able to both carry out the correct action, and make no eye movement when decoys were presented.

To test their action knowledge, we compared the probability of selecting the correct dot in all trials where an action was carried out, and compared this to a chance level of 0.14 (out of six possible actions or no

response). Participants were able to perform the correct action at a greater-than-chance level, both on the first day (in the lab using a desk-mounted tracker; mean: 0.572; wilcoxon(49)=1431, p<0.001)) and the second day (in the fMRI scanner using a mirrored eyetracker; mean: 0.56; wilcoxon(49)=1275, p<0.001; Figure 5b).

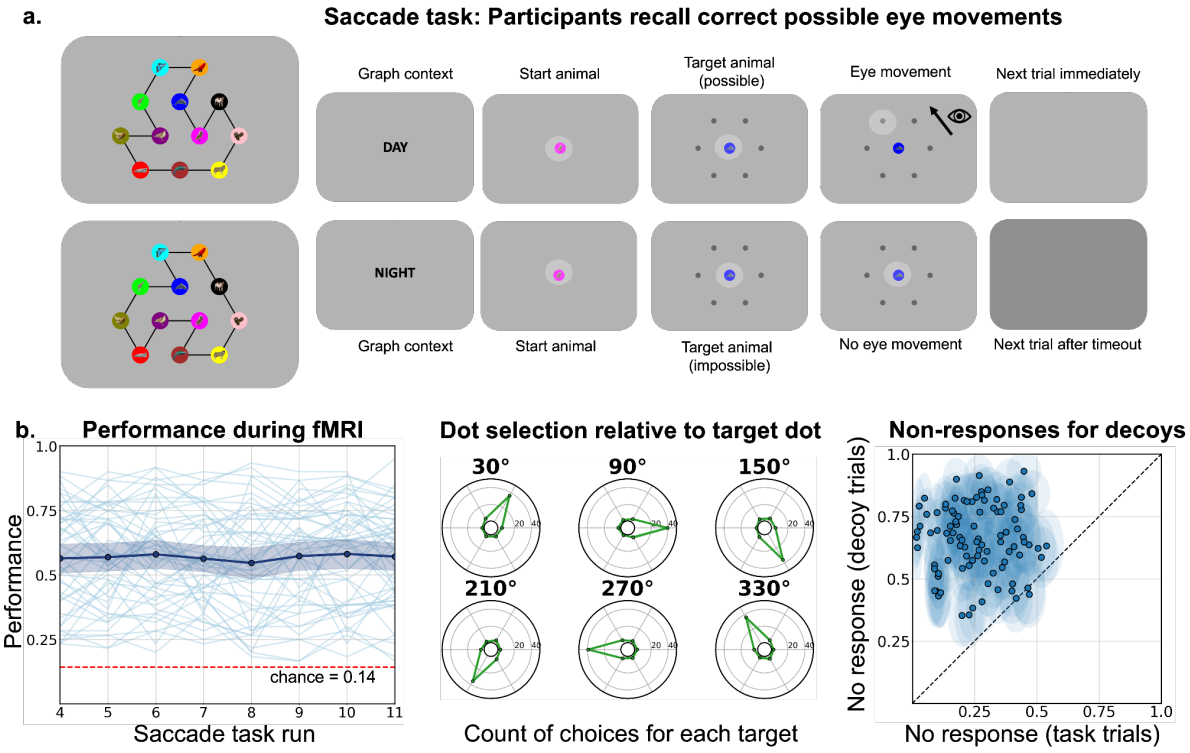

Figure 5: Participants selected the correct action from each animal image according to graph context*. a. Trial timeline (not including inter-stimulus intervals between presentations; see Methods for details). Participants were asked to view a 'start' animal image in isolation, and imagine possible successors. They were then presented with a 'target' animal, upon which they had to either (1) select a dot in the relative direction of this animal compared to the 'start' animal using an eye movement (task trial), or (2) if the target was not possible within that graph, keep fixation on the centre (decoy trial). b left: Performance in the fMRI scanner during functional runs. b centre: Count of dot selection, divided by target direction: participants mostly selected the correct dot direction in task trials in which a response was registered.2b right: Participants correctly identified decoys, selecting dots less in decoy trials than task trials.*

To further explore this, we tested if performance varied as a function of position or graph identity using a two-way repeated-measures ANOVA using graph and position as factors. This was not the case (graph: F(1,576)=0.05, p=0.82; position: F(5,576)=0.16, p=0.98; interaction: F(5,576)=0.13, p=0.99; see Figure

2b centre). Reaction time was also not influenced by graph or position for conditions of interest, suggesting no relevant differences in task difficulty between conditions (graph: $F(1,14972)=0.02$, $p=0.63$; position: $F(5,14972)=0.34$, $p=0.64$; interaction: $F(5,14972)=0.78$, $p=0.17$).

However, these correct responses could have theoretically been carried out by relying on positional knowledge of the animals (e.g. the lion is always to the right of the dog). To ensure that participants were actively recalling the graph context, we checked that participants were less likely to respond for decoy than non-decoy trial, as measured using dprime scores (hits were taken as no response during decoy trials; false alarms were no response for task trials). This was indeed the case: all participants were more likely to respond when the target animal was a possible transition within the cued graph ($wilcoxon(49)=1225$, $p<0.001$, see Figure 5b right).

3.4.3 Visuospatial action representations in the entorhinal cortex

We predicted that, in line with our previous findings, the entorhinal cortex would represent the learned actions (eye movements) associated with each animal image. If these representations generalise, this should apply across all states, regardless of graph context.

As with the previous experiment, each state afforded two actions. Each action corresponded to a directional eye movement. Therefore, a given image might afford eye movements up-left and right in one graph (i.e. day or night), but up-right and down-left in the other graph. This allowed us to compare images based on the similarity of actions across all image-graph combinations: if two images both afforded the same eye movements in their given context, we would predict similar neural patterns (Figure 6a).

In practice, we compared neural pattern similarities in each hemisphere to two models of action-based state similarity (Figure 6b). One predicted state similarity in terms of the minimum spatial transformation required to map between the actions of each state (i.e. how far apart actions are) using cosine distance between action pairs, while the second predicted state similarity on the basis of discrete action identity (i.e. how many actions are shared between states). Whereas in our previous experiment these would have produced identical model matrices, here the projection on visual space allowed us to test models based both on action identity and on action distance (in terms of angle). Models of link distance and Euclidean distance between states (during learning) were controlled for in a partial correlation, thereby accounting for effects

related to screen position or transition probability during the learning phase of the task. For details, see Methods.

We found that the right entorhinal cortex represented the action possibilities for each state in terms of minimum action transformation (action distance model: t(48)=2.49, p=0.0082; action identity model: t(48)=-1.31, p=0.90; one-tailed one-sample t-test against 0; Bonferroni-corrected alpha level of 0.0125; see methods). As with the previous experiment, we found no evidence for an action representation in the left hemisphere (action distance model: t(48)=-0.81, p=0.788; action identity model: t(48)=0.07, p=0.472).

As before, it should be noted that this result could be part of a wider activation or set of regions. In future analyses, a whole-brain searchlight will be run to understand the wider distribution of action representation. As an intermediate control, however, we also ran the same analyses for action distance in the wider medial temporal lobes, using the same ROI as the previous experiment (hippocampus and subfields, parahippocampal gyrus, entorhinal cortex). We found no significant effect of action distance in the wider MTL, either looking bilaterally (t(48)=0.92, p=0.180) or specifically at the right hemisphere (t(48)=1.47, p=0.074). This may be taken as an early indication that our results are indeed specific to the right entorhinal cortex, but further analysis and independent testing of each region in the MTL will be needed to fully establish this.

These results replicate our previous finding that the right entorhinal cortex represents states in terms of action similarity. The prevalence of the action distance model may possibly indicate a more spatial framework (as our model relies on a continuous distance measure calculated from the orientation of different actions), but this could simply be a result of the visuospatial projection of our task and would require further testing to draw conclusions about the general nature of action representation.

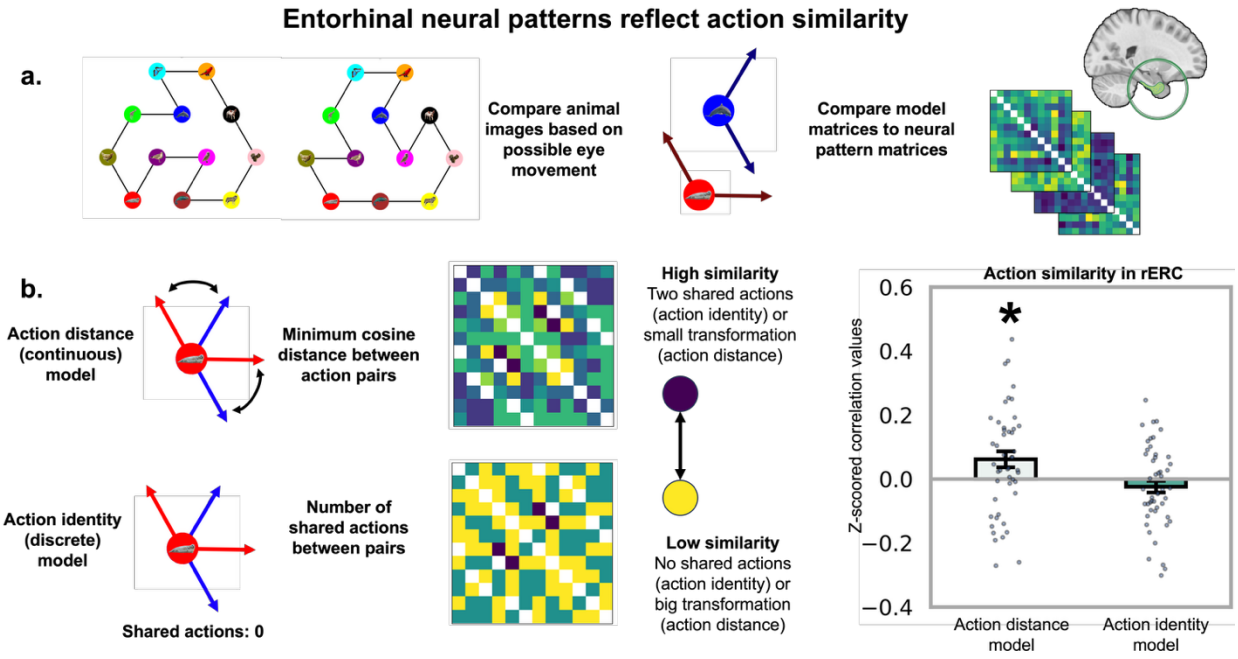

Figure 6: Representation of action in the entorhinal cortex. *a. Model RDMs were constructed based on possible eye movements from each animal picture (state), and then compared to neural RDMs using tie-correction spearman correlation. b left: model RDMs were constructed based on the minimal distance between action pairs using cosine distance or on the number of shared actions across images. The former (action distance, above) predicted that states would be more similar if their afforded action pairs were more similar in terms of angle. The latter (action identity, below) predicted that states would be more similar if afforded action pairs contained the same actions. The resulting matrices were therefore continuous (action distance) or count-based (identity). b right: The action distance model was represented in the right entorhinal cortex: correlation values between the neural RDM and model RDM were significantly greater than 0 in a one-tailed, one-sample t-test at a group level.*

3.4.4 Representation of affordances within and across graph contexts

Having replicated the finding that the right entorhinal cortex represents states in terms of action similarities, we tried to answer our main question: are representations of action context-specific, or do they generalise across different task contexts?

To do so, we divided our action distance model matrix into two sub-matrices, which tested for action-based state similarities either within or across the same graph context. While the previous finding provided

evidence for a general representation of action information , it was nonetheless possible that this could be driven more strongly by a within-graph representation than an across-graph representation.

We found a representation of continuous action similarities within graph ($t(48) = 2.51$, $p=0.0078$, one-tailed one-sample t-test against 0; Bonferroni-corrected alpha level of 0.025), and a non-significant trend across graphs ($t(48) = 1.70$, $p=0.048$). In the case of context-specific representations, we would expect the within-graph representation to be stronger than the cross-graph representation. Therefore, we compared the Fisher Z transformed correlation coefficients resulting from the model comparisons using a paired t-test (i.e. we predicted that correlation values with the within-graph RDM would be different to correlation values with the cross-graph RDM) but did not find a significant difference ($t(48)=1.07$, $p=0.290$, two-tailed paired t-test; alpha level of 0.05). Therefore, we found no clear evidence for an 'entanglement', whereby action representations would be specific to a within-graph context. In short, the result was somewhat inconclusive: while we found no difference in the action representation between the two contexts (which might have potentially hinted at a generalisable action representation), the cross-graph action representation was not significant.

3.4.5 Rotation of affordances across graph

As the representation of action across graphs was unclear, we asked if this could be explained by a change in the alignment between the two contexts. As I will briefly discuss, experimental work has suggested that representations of relational structure may rotate uniformly across task contexts. This may allow both a generalisation of scalar information such as distances, and a differentiation of contexts. Therefore, we asked if similar coding principles may be applied in our task.

In the rodent literature, it has been widely observed that entorhinal representations generalize metric information through rotational realignment (Hafting et al., 2005; Fyhn et al., 2007; Høydal et al., 2019). That is, the orientation of neural populations in the entorhinal cortex and surrounding regions (specifically object-vector cells, grid cells and head direction cells) rotate in sync when an animal is exposed to a new environment. Object-vector cells, which have been considered a possible substrate for action representation in the entorhinal cortex (Whittington et al., 2020), are paradigmatic for our experiment: If in a given environment two object vector cells, OVC-A and OVC-B, fire at 15 degrees to an object and 90 degrees, respectively, in a different environment both vectors might rotate equally, for example to 45 and 120 degrees. In short, the cells often maintain relative angular distance but change their absolute vectorial direction by a common rotation, typically in line with environmental axes (Krupic et al., 2015; Høydal et

al., 2019; Andersson et al., 2021). Recent work, in fact, has observed rotations of relational information in primate entorhinal cortex (Veselic et al., 2025) and human parietal cortices (Yu et al., 2025). Although these prior analyses looked at rotation of grid-like signal, we reasoned that similar coding principles might be applied to our task.

Based on this, we asked if the change of task context could have elicited a similar rotation of neural representations in the right entorhinal cortex. To test this, we ran models of state similarity in which all the actions of the first graph were rotated before comparison to the other graph (via minimal cosine distance, as before). This can be seen in Figure 7b. For example, suppose that we started with an action pair of 0 degrees and 180 degrees (taking only two actions for simplicity). These would be considered maximally dissimilar, with a cosine distance of 2 (via the calculation 1-cos(0-180)). A rotation of 90 degrees clockwise, however, would adjust the first action to 90 degrees without changing the second action. Therefore, the cosine distance would become 1-cos(90-180), which is 1. This transformation would then be applied uniformly, adjusting all cross-graph distances and recalculating the entire model RDM accordingly. This 90-degree model could be used to test the hypothesis that action orientation is rotated by 90 degrees across contexts.

In our case, we had no prediction about the exact rotation. Therefore, we aimed to generate a series of models for many possible rotations and test if a specific rotation angle produced model RDMs which correlated well with neural RDMs (Figure 7b centre). This generated a problem of multiple comparisons: by generating and testing a lot of models, we amplify the risk of finding a false positive. Indeed, in 360 different rotation models it would be very likely to find a false positive unless stringent correction were applied! To account for this, we drew on the fact that adjacent models were highly correlated and would therefore form clusters of angles where the rotated action RDMs would correlate with entorhinal RDMs. This enabled us to run cluster correction, following standard procedures used for analysing time series.

Therefore, we generated rotated models for every integer degree of rotation (i.e. 0-359, step size of 1) and correlated each model with each subject's neural RDM from the right entorhinal cortex. The resultant correlation values were then Fisher Z transformed and tested against 0 using a one-tailed t-test to generate a series of t-values. Clusters of adjacent angles were identified using a threshold of 1.677 (corresponding to an alpha value of 0.05 with 48 degrees of freedom), before being compared to a random distribution of cluster masses based on sign flipping. This method identified a significant cross-graph effect of action affordances at a cluster of 91-194 degrees (cluster permutation; df(48), p=0.0191, alpha=0.05; Figure 7b right). This preliminary result indicates that action representations rotate across contexts, potentially hinting

at a coding scheme whereby rotation signals a change in context, but the same representational format is used across graphs. Nevertheless, it should be highlighted that further analyses will be necessary to understand this better, both by looking at whole-brain representations of action rotation and the interaction between rotation and eye movement behaviour.

In sum, the current results provide a replication of our previous finding that the right entorhinal cortex represents possible actions in a relational structure. While we were unable to find clear evidence for an entangled or factorised action representation, a potential mechanism for both generalisation and context-specificity was identified by looking at the relative rotation of action information across contexts. As will be discussed in further detail below, however, further analyses will be necessary to confirm this result and understand how rotation interplays with wider brain networks and eye movement.

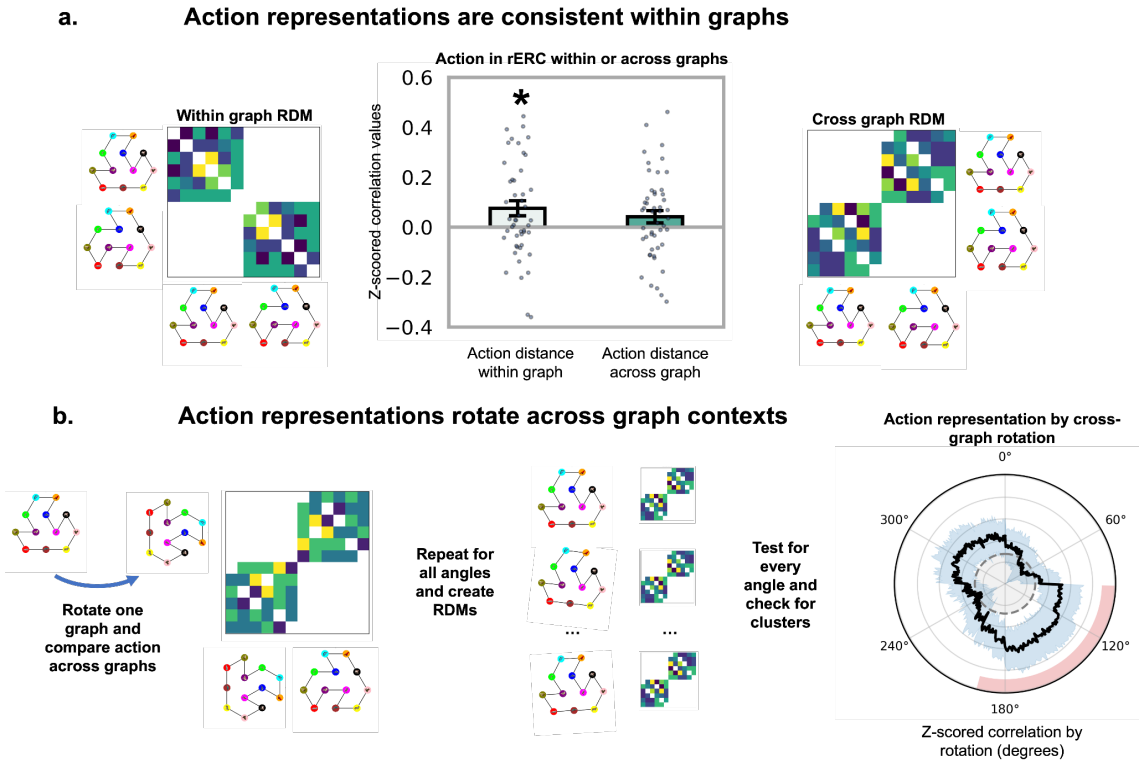

Figure 7: Representation of action across or within graphs. *a. Model RDMs were tested both within or across graph structures. Action distance was represented within the same graph. b left: we tested if action representations rotated across graphs by comparing action distance when one graph was rotated relative to the other. This created 360 models, one for each angle, which were tested against 0 using a cluster*

correction. b right: we found a cluster around 91-194 degrees, indicating that neural representations of action rotate across graphs.

3.5 Discussion

Relating entorhinal representations to behaviour has been an enduring challenge since the discovery of grid cells (Hafting et al., 2005) and subsequent theoretical accounts predicting that the entorhinal cortex maps out the relational structure of tasks (Whittington et al., 2020), despite hints that entorhinal representations are responsive to learned patterns of behaviour both in rodents and humans (Stachenfeld et al., 2017; Jeffery, 2021; Ginosar et al., 2023). In the previous experiment, we demonstrated that the entorhinal cortex represents possible actions in a conceptual space. Here, we repeat this finding by showing that actions are represented in the right entorhinal cortex, and show that across graph contexts actions are represented at a uniform rotation of approximately 90 to 180 degrees.

The main question this experiment set out to answer was whether action representations are context-specific or generalise across different task contexts (Whittington et al., 2022). Our previous experiment assumed that action was consistent across different numbers, but we only presented a single task context. Nonetheless, generalisation of action is an important prediction of theoretical accounts suggesting the entorhinal cortex provides a 'basis set' of representations which can be used for compositional generalisation (Bakermans et al., 2025): to recompose relational structures in aid of flexible behaviour, we would expect factorised 'building blocks' of sensory stimuli, action and position. For example, the same state may imply a different action in the case of changing reward. Interestingly, a recent review argued that the level of factorisation may depend on task demands: the more experience you have with a given set of actions across different contexts, the more likely you may be to (neurally) separate action from other features of the relational structure (Whittington et al., 2022). We designed our task to highlight action similarities across contexts, using the same oculomotor responses and task design. Nonetheless, an interesting future direction may be to understand how action representations change throughout a longer learning period.

In a targeted ROI-based approach, we first found that the right entorhinal cortex represents action similarities. This replicates the results from the previous experiment in a different domain, lending weight to the account that the entorhinal cortex represents task structure across different domains. The action representation we find is based on the minimal cosine distance between action pairs, which can be interpreted as the minimum possible transformation between actions. This may be driven by the visuospatial nature of our task: more similar actions in terms of angle correspond to more similar oculomotor movements (i.e. engagement of the same muscles) and more similar visual features during learning (e.g. closer relative target position). In general, it might be argued that distance-based similarities require a reference frame

which may be considered more ‘map-like’ (Peer et al., 2021) as these similarities would not scale with angle in a graph-like representation (e.g. Garvert et al., 2017). Indeed, previous fMRI work has shown that the entorhinal cortex generalises better for relational structures which can be mapped onto a 2D space (Mark et al., 2024), although this result does not provide strong evidence for a map-like representation per se. While these results appear to align with conceptual accounts which suggest entorhinal representations constitute a ‘spatial scaffold’ for cognition (Chandra et al., 2025), the current experiment was not designed to test this hypothesis directly and so further work will be needed to understand if this distance-based action representation emerges as a feature of our task or is a more general feature of entorhinal representations of relational structure.

In terms of generalisation, our results do not clearly indicate whether our main effect is driven by within-graph or cross-graph action similarities. When our main RDM is broken down into within- and cross-graph RDMs we do not find evidence for a cross-graph representation of action, but also not a significantly stronger representation of action within graph. Future analyses might seek to understand this result better by looking at individual differences. It might be the case that some participants generalise action information without rotation, whereas others rotate. More generally, it would be important to run further analyses at a whole-brain level to understand these results better and be sure that our models are accurately capturing action similarities. Regions which respond to oculomotor action planning, such as the posterior parietal cortex, should be expected to represent action similarities both within and across graphs. Additionally, as rotation has previously been observed mostly in the entorhinal cortex and adjacent regions, we may expect to find rotated action representations specifically in the medial temporal lobes.

The rotation of the action map across contexts mirrors previous results in rodents and primates. In rodents, object vector cells, grid cells and head direction cell representations rotate in tandem (Høydal et al., 2019), often aligned to the axes of new environments (Krupic et al., 2015). While it is unclear if an axis from the experimental setup might explain the specific rotation we see (e.g. rectangular screen), it is possible that the shape of the ‘zoo’ might make some axes more salient than others. As far as we are aware, this is the first evidence of task context information inducing rotation in the entorhinal cortex in humans, although previous work has demonstrated rotation in grid-like signal across different stimulus sets in non-human primates (Veselic et al., 2025) and in the human parietal cortex (Yu et al., 2025). Given that the other task features stay the same, and indeed the same input is provided from the oculomotor system during learning, this may be interpreted as a functional shift - rotation may help to distinguish task contexts. Future analyses will aim to understand if more rotation helps participants to perform better in trials where the context changes.

An interesting question regards how eye movement might interact with rotation. Eye movements do not rotate across contexts, as actions correspond to a specific direction on the screen. In our previous experiment, we found a potential link between eye movement behaviour and neural representations of action. The current experiment complicates the picture: eye movements cannot both reflect possible action during the initial image presentation *and* rotate across contexts. Therefore, whereas in the previous chapter we suggested a direct input from oculomotor systems to the medial temporal lobes, to explain the current results one might invoke a mediating mechanism which translates eye movement signals into the correct format. This is somewhat surprising given the obvious screen reference coordinates: the simplest solution would involve no rotation. Given this, it could be argued that rotation is only an efficient coding strategy if there is some computational benefit. In this case, it could be used for disentangling task contexts. If this is the case, one graph (perhaps the best-learned) may reflect eye movements in screen coordinates, while the other would rotate. That said, our previous study showed only marginal evidence for a link between eye movement and neural responses. Therefore, it is also possible that eye movements are not recruited during the recall phase of the task, despite the visuospatial framework. This remains to be tested.

A related analysis may seek to understand how behaviour influences neural representations, both in the entorhinal cortex and the wider brain. As mentioned previously, entorhinal representations appear sensitive to patterns of learned behaviour or statistical regularities (Baram et al., 2021). During the learning phase of our task, participants were free to explore the graph as they liked (despite being asked to try to alternate directions). This means that some participants displayed a tendency toward following the graph in one direction (e.g. anticlockwise). Therefore, different states may be different not only in terms of action possibilities, but in terms of how likely participants were to select a given action. There are multiple information-theoretic measures here which may help us to understand how neural patterns relate to this behaviour. For a start, we could use Kullblack-Leibler Divergence to quantify the difference in action choice distributions between states in order to calculate behaviourally modulated action RDMs. During the saccade task, we may expect neural responses related to surprisal in brain regions involved in prediction (a widely-used paradigm in language or visual processing, e.g. Heilbron et al., 2022). In terms of the medial temporal lobes specifically, Successor Representation models predict that the hippocampus represents a discounted graph of state occupancy, while the entorhinal cortex represents a low-dimensional projection of the former (Stachenfeld et al, 2016). These may be calculated for each participant and symmetrised before being compared to neural representations in the hippocampus and entorhinal cortex using the same ROI-based RSA approach as before.

A final note should regard other brain regions, which were briefly mentioned above. For the purpose of this thesis, we have limited our analyses to the entorhinal cortex. Nonetheless, the oculomotor system should be involved in our task! We would expect the parietal cortex to represent actions both within and across contexts as part of motor planning (Whitlock et al., 2012; Nitz, 2014; Vericel et al., 2024), perhaps in a count-based form more similar to the representation of specific efference copies, and actual eye movement should elicit responses in both visual regions and frontal eye fields. Moreover, if the results observed here correspond to a re-use of spatial machinery in a visual domain (Buzsáki & Moser, 2013; Bellmund et al., 2018; Bottini & Doeller, 2020; Chandra et al., 2025), we might expect to see a transformation of predicted eye movement from an egocentric form to an allocentric form via the retrosplenial cortex (Bicanski & Burgess, 2018).

In sum, our results provide further evidence that the human entorhinal cortex represents the action component of relational structures. We find no clear evidence for a context-specific or context-general representation of action. However, we observe a ‘rotated’ representation of action across contexts in the right entorhinal cortex, such that action pairs are considered more similar following a rotation of 90-194 degrees across contexts. While this result will require further investigation to understand how eye movements interact with rotation, it may hint at a mechanism for relational memory which maintains the same representational format across contexts but indicates context-specificity via rotation.

Conclusion

Higher-level mental representations are sometimes thought of as disconnected from behaviour. In cognitive science, the raw behaviourism of Skinner and disciples (Skinner, 1957) is often contrasted with the abstract symbolism of Chomsky (Chomsky, 1959) and Fodor (Fodor, 1975). Cognitive maps were imagined as an organising principle for all relational thought, well beyond spatial navigation (Tolman, 1948), and indeed have been duly criticised on the basis that they do not easily fit into the pathway from perception to action (Gibson, 1979). However, despite their domain-generality (Tolman, 1948) and apparent compositionality (Frankland & Greene, 2020; Bakermans et al., 2025), even the most abstract, seemingly-invariant representations in the entorhinal cortex appear to adapt to the task at hand. The evidence presented in this thesis suggests that entorhinal cortex neural populations are not just modulated by environmental features, but represent the actions agents can carry out at any given moment. As entorhinal representations of relational information are considered building blocks for cognitive maps (McNaughton et al., 2006; Whittington et al., 2022), this might start to reintegrate cognitive maps into the perception-action loop used for influential theories with broad explanatory value such as predictive coding (Friston & Kiebel, 2009; Rao & Ballard, 1999).

The unifying question behind this thesis has been to ask if entorhinal representations of relational structure contain action representations across domains. This appears to be the case. In an experiment where participants learned to associate numbers with numerical operations, recall of these same numbers induced BOLD response patterns in the right entorhinal cortex which were modulated by action-based similarity (Chapter 2). A similar paradigm in visual space elicited neural patterns which reflected the minimum transformation distance between action pairs (Chapter 3). In our experimental designs, we ensured that action itself was being targeted by isolating possible transitions from position. In the first experiment, actions were repeated across repeating graph modules such that the distance between numbers or states could not explain neural representations. In the second experiment, animal images were associated with different positions during training and testing, and subsequently presented with a spatial jitter during the task in the fMRI scanner. In short, we find representations of possible actions in the entorhinal cortex in two separate experiments. One experiment drew on a repeating graph structure with numerically labelled states, and the other involved two visuospatial graphs learned through visual exploration. In both cases, the affordance representation was limited to the right hemisphere and could not be explained by other features of the learned maps such as the distance between states.

How does this fit with prior work? Previously, we have defined action as a domain-general transition between states. This is a wide definition which therefore encompasses not only internal action (Dayan et al., 2012) but also motor action (or chains of motor actions, such as in navigation), insofar as this elicits a change between states. Therefore, if the entorhinal cortex learns state-action graphs as previously suggested, we might expect impairments in action learning across domains in cases of damage to the entorhinal cortex. As noted previously (Section 1.3), focal damage to the entorhinal cortex is rare, and so much prior evidence comes either from wider damage to the medial temporal lobes, or early Alzheimer's disease. However, neuropsychology studies based on medial temporal lobe damage have typically not observed a pronounced motor deficit on the level of memory impairments: famously, H.M. was able to acquire procedural motor skills even following his operation (Squire, 2009). Nonetheless, amnesiacs do show a generally lower motor performance in most tasks (Shadmehr et al., 1998), and it has been observed that an amnesiac with extensive damage to the hippocampal formation (including the bilateral entorhinal cortex) showed specific deficits in flexible motor action selection and failed to learn a task which required action-association learning (McDougle et al., 2002). Some specificity to the entorhinal cortex may be suggested by primate evidence: rhesus monkeys are impaired in visuomotor association learning only in the presence of entorhinal lesions, not hippocampal lesions (Yang et al., 2014). Neuroimaging evidence has also shown that learning the same motor action associations results in more similar neural patterns in the entorhinal cortex, but not the hippocampus (Hindy and Turk-Browne, 2015). Therefore, it seems as though the entorhinal cortex may play some role in how motor actions are learned and retrieved. This would be in line with the proposed role mentioned in Section 1.3, whereby the entorhinal cortex carries out association learning in support of relational memory (Klingberg et al., 1994; Buckmaster et al., 2004; Garcia and Buffalo, 2020). On the other hand, there does appear to be an imbalance between the impact of MTL damage during memory tasks (full retrograde and anterograde amnesia), navigation tasks (path selection deficits) and motor tasks (limited but not catastrophic impairment, mostly related to memory-related tasks). This may be seen as a challenge to the account that the entorhinal cortex is a truly domain-general region for action learning. Nonetheless, this may also reflect the relative importance of learning relational structures for different tasks, as both navigation and motor tasks can often be solved through simpler cue-based strategies or existing knowledge that does not require learning new associations. Although our results extend action representations beyond the motor domain, our findings may be a function of the task setting (which rely on teaching a graph structure through associative learning) rather than a general feature of any conceptual or visual task.

Various questions remain about our findings. As a start, we have no clear explanation for why both studies find the right entorhinal cortex, and not the left. Although there have been proposals on lateralisation in the hippocampal-entorhinal system suggesting more spatial information in the right hemisphere (O'Keefe & Nadel, 1978), and fMRI grid-like signal was originally observed only in the right hemisphere (Doeller et al., 2010), other studies have found only the left hemisphere (Sigismondi et al., 2024). Therefore, further study would be needed, ideally with more noise-resilient methods such as intracranial recordings, to understand why our effects are lateralised to the right hemisphere.

In the second study we tried to answer if action representations were context-specific or context-general. As with the previous experiment, we first found that the right entorhinal neural similarities between states correlated with a matrix of state similarities based on action similarity. We then tested within and across contexts separately using subsets of the main RDM. Although we did not find evidence for cross-context action representations, the action representation within context was not significantly stronger. Further work will be needed to understand this result better. In general, this reflects a slightly wider question about the nature of action representations in the context of relational structures: do they reflect a context-specific distortion, or a separate substrate which may be flexibly recomposed across different states and structures? Whereas rodent work has largely focused on distortions as a result of environmental changes, which could be explained in terms of action constraints within a given context (Jeffery, 2020; Ginosar et al., 2023; Dorrell et al., 2022), some modelling work has described action as a flexible component which can apply across different stimulus sets (Whittington et al., 2020; Bakermans et al., 2025). Human neuroimaging work has shown both that the same reward probability distributions elicit more similar entorhinal neural patterns across contexts (Baram et al., 2021) and that the same motor action associations are coded similarly in the entorhinal cortex across different stimuli (Hindy and Turk-Browne, 2016). This may hint at coding principles which apply the same relational structure across contexts. Nonetheless, this has not been directly addressed targeting action across different task contexts. In our experiments, we find an action representation which is consistent across states and within contexts in both experiments. In the second experiment, it appeared as though action representations span across all states independently of task context, but when we split the RDMs into within and cross-graph models we did not find evidence for a cross-graph representation. Therefore, it remains unclear if action is represented similarly across contexts. To test this further, we may look at representation of action during the response period (i.e. do the same eye movements elicit similar entorhinal signals across contexts?), or probe the nature of the underlying relational structure by comparing non-action variables across graphs, such as context-specificity of distance information or link distance relative to action.

It should also be noted that the functional role of action representations remains an open question. In the work presented here, we find no clear link between neural representations of action and behaviour (although eye movement lateralisation does predict performance). This lack of evidence is hard to interpret but may be driven by overtraining (in Chapter 2) or noisy eyetracking data from the MRI scanner (in Chapter 3). Theoretical work provides predictions on how entorhinal action representations are linked to behaviour, although it should be noted that most of these theoretical accounts aim to explain grid cells rather than the entorhinal system as a whole, which may not follow the same computational rules as the grid system. They may reflect a bias in the system during learning of 'actionable' representations (Dorrell et al., 2022; Ginosar et al., 2023), or provide a constraint on future action via predictive or enactive representations (Stachenfeld et al., 2017; Jeffery, 2021). These may be tested in the second experiment by testing if entorhinal representations are modulated by action probability distribution during learning (according to the actionability hypothesis) or predict future eye movement (predictive representation hypothesis). It should be noted, however, that these are not mutually exclusive hypotheses (indeed, one might expect that learning patterns predict behaviour) and the link between entorhinal representations and behaviour may be difficult to establish due to the complexity of the processes involved and multiple possible strategies (Ekstrom et al., 2020; for a discussion in terms of reinforcement learning models see Gershman & Daw, 2017).

In the second experiment, we find that an RDM of rotated action similarities predicts entorhinal neural pattern similarities. This suggests that the identified action representations rotate across contexts, mirroring cross-environment or context rotations in entorhinal cell populations (Hafting et al., 2005; Fyhn et al., 2007; Høydal et al., 2019) and grid-like signal in non-human primates (Veselic et al., 2025) and humans (Yu et al., 2025) – though it should be noted that rodent navigation is a very different task setting from human MRI, so direct cross-species comparisons are difficult to draw. Nonetheless, the general computational trait of maintaining the same structural representations but rotating across contexts is in line with the proposed role of the entorhinal cortex as a generalisation machine (McNaughton et al., 2006; Whittington et al., 2020). Our result suggests that the same representational format for actions is used across contexts, enabling us to compare actions between the two graphs (post-rotation). This might indicate re-use of the same structural features (as per Bakermans et al., 2025) across different contexts. It might be observed that rotation appears to require a spatial reference frame to be meaningful, which could align with theories about the use of entorhinal cortex as a spatial scaffold (Chandra et al., 2025), but this may be driven by the visuospatial task framework, as noted in the discussion of the previous chapter. An interesting future step would therefore be to test this in conceptual spaces with less spatial grounding, possibly even in higher dimensions where the axis of rotation is less clear than the current 2D visuospatial projection. In abstract spaces, it has been suggested that a combination of lower-dimensional spaces may be used to build up to

higher-dimensional representations (Iyer et al., 2024); in this case, we may expect rotations only in dimensions where task context changes, in line with results suggesting that the hippocampal formation carries out dimensionality selection or expansion of key task variables (Theves et al., 2020; Boorman et al., 2021; Kerrén et al., 2025).

A related question applies to how participants encode and retrieve action information. Although the second experiment asked participants to learn a visuospatial map, retrieval of the correct map required learning the correct animal-position-graph association. This means that in order to solve a given trial, participants first had to retrieve the correct position from memory using the animal or colour cue. This may mean that other semantic information is encoded in the task - although a simple stimulus-response mapping would have been possible, many participants reported active engagement with the animal identity or colour (even using animal names) in order to solve the task. If this information is indeed represented, it may help our understanding of how feature selection interacts with eye movement and rotation, as mentioned above. As animal images were drawn from a dataset with ratings in a high-dimensional space, a possible future step may be to see if hippocampal neural patterns or eye movement predict semantic similarity during retrieval (for hippocampal signatures of semantic retrieval in intracranial EEG during a foraging task, see Solomon et al., 2019; and Viganò, Giari et al., 2025; for semantic distance in eye movement while retrieving animal names see Viganò et al., 2024) - and if this is integrated into a multi-feature representation which also encodes other task variables such as colour, graph identity, position and action (for hippocampal cells which encode multiple task or mnemonic features in non-human primates and humans, see Bernardi et al., 2021, Kolibius et al., 2023 and Courellis et al., 2024).

In future work, it will be important to understand how action representations are related to eye movements. In both experiments, eye movements play a key role. In the first experiment (Chapter 2), the task elicits lateralised eye movements in the direction of afforded actions. This goes in line with previous work suggesting automatic recruitment of visuomotor systems during arithmetic, expressed both in behaviour (Dehaene et al., 1993; Loetscher et al., 2008) and neural signals (Walsh, 2003; Summerfield et al., 2020). Interestingly, the strength of the gaze lateralisation effect predicts performance during training, and appear to be connected to neural representations in the entorhinal cortex, hinting at a functional role in our task. The connection to entorhinal representations is novel but will require replication due to the borderline effect (which could be partially due to noise in either the neural data itself or the implementation of DeepMReye, which was run without fine-tuning due to the last of available data). In the second experiment (Chapter 3), eye movement is the bedrock of the task, which relies on learning eye movement associations during visual exploration. In both cases, therefore, action representations appear to be tied to eye movement behaviour.

As mentioned in Chapter 2, the findings in the first study may hint at an embodiment of non-spatial velocity input in eye movement. If cognitive maps are based around a two-dimensional continuous space as suggested by O'Keefe and Nadel (1978), the format of velocity input should match this; eye movements could provide a useful constraint, as they are naturally mapped to a low-dimensional frame. However, we did not directly test this, and the format of medial temporal lobe structure representations is contested (Cohen and Eichenbaum, 1991; Peer et al., 2021). Nonetheless, movements in conceptual spaces are occasionally characterised as movements of attention (Viganò et al., 2024) which are typically expressed in eye movement behaviour (Groner & Groner, 1989) and could function as an action input regardless of representational format. If this is the case, we should find that task-related eye movements drive changes in entorhinal representations. A further intriguing possibility on this basis is that previous results in the literature which did not use eyetracking may also have been accompanied by unobserved shifts in visual attention. This could be retrospectively tested using DeepMReye, although as noted our own results were noisy despite the clear spatial mapping and strong eye movement effect observed on the first day. Nonetheless, it is important to note that both our experiments rely on paradigms which are both naturally projected onto two-dimensional spaces, and would both be expected to elicit eye movements. Therefore, it remains to be seen if eye movements are recruited in other domains with less of a clear link to eye movement (e.g. discrete semantic spaces; Viganò et al., 2020), although there are some early hints that this is indeed the case for other domains such as colour or word frequency (Viganò et al., 2024), and if the resulting eye movements can also be linked to neural representations of task structure.

The role of other brain regions beyond the entorhinal cortex remains unclear. In our first experiment, a whole-brain searchlight analysis was inconclusive. An exploratory analysis revealed a posterior parietal representation of affordances. This provides an interesting parallel with spatial navigation, in which the posterior parietal cortex is proposed as a region which gathers self-motion signals as part of an egocentric spatial representation (Avila et al., 2019; Alexander et al., 2022; Vericel et al., 2024) prior to conversation into an allocentric framework in the retrosplenial cortex (Bicanski & Burgess, 2018). Other work has suggested that the parietal cortex organises information into a low-dimensional space (Summerfield et al., 2020) and plans routes through mazes (Andersen & Buneo, 2002; Crowe et al., 2004; Alexander et al., 2022; Nitz, 2014).In light of this, it should be noted that the exact role of the parietal cortex in our task is unclear, due to the multiple proposed functions. The second experiment may be better-placed to distinguish parietal and entorhinal representations, both due to multiple models of action representation and access to eye movement data, which may help to disambiguate neural signal related to eye movement activity from neural activity related to abstract features of the task. Finally, other involved regions may include frontal and striatal areas, which have an established role in action planning (Miller & Cohen, 2001; Cox & Witten,

2019). Of particular note is the prefrontal cortex: a recent rodent study discovered modules of task-progress cells which were linked to spatial locations, thereby allowing ‘automatic’ task resolution by following these attractors (El-Gaby et al., 2024). A model based on these findings demonstrated that velocity inputs can be calculated from these cells, which would work as a potential alternative mechanism to the ocular mechanism proposed above (Whittington et al., 2025). While we are unable to measure single neural responses (or, indeed, any neural responses at the timescales required to identify task-progress cells), the prefrontal cortex may play a role in calculating actions.

Finally, it should be noted that a large body of fascinating findings have not been taken into consideration in the current work, predominantly due to the low temporal resolution of fMRI. These include neural findings which have been proposed as a link to behaviour, such as replay (Ólafsdóttir et al., 2018; Nour et al., 2021; Bakermans et al., 2025) or phase precession (Qasim et al., 2021), as well as physiological features of entorhinal cells such as theta oscillations in grid cells (Buzsáki & Moser, 2013) or more recent, less well-understood discoveries which nonetheless might link to behaviour such as alternating exploratory movements in estimated position from grid cells (Vollan et al., 2025). While these are beyond the scope of the current work, it is possible that these fast dynamics could be instrumental in rapid exploration of relational structures to establish which actions may be carried out at any given point.

The current work has aimed to elucidate how entorhinal representations of relational structure are tied to behaviour, through the lens of action. It is shown in two separate eyetracking and fMRI experiments that neural representations in the entorhinal cortex are sensitive to the learned action structure of a relational structure. This was the case both in the visuospatial domain, as well as a conceptual task using arithmetic. In the first experiment we draw a link between eye movement and hippocampal-entorhinal cognitive maps. In the second experiment we provide evidence that these action representations rotate across different graphs, hinting at a rotation-based mechanism for separation of task context. Nonetheless, important questions remain about whether action is integrated into entorhinal representations or separable; how and why eye movement is connected to movement in relational structures; how other brain regions interact with action representations, and, most crucially, to what extent action representations guide behaviour.

Bibliography

Abraham, Alexandre, Fabian Pedregosa, Michael Eickenberg, Philippe Gervais, Andreas Mueller, Jean Kossaifi, Alexandre Gramfort, Bertrand Thirion, and Gael Varoquaux. 2014. “Machine Learning for Neuroimaging with Scikit-Learn.” Frontiers in Neuroinformatics 8. https://doi.org/10.3389/fninf.2014.00014.

Agarwal, A., Sarel, A., Derdikman, D., Ulanovsky, N., & Gutfreund, Y. (2023). Spatial coding in the hippocampus and hyperpallium of flying owls. *Proceedings of the National Academy of Sciences*, *120*(5), e2212418120. https://doi.org/10.1073/pnas.2212418120

Alexander, A. S., Tung, J. C., Chapman, G. W., Conner, A. M., Shelley, L. E., Hasselmo, M. E., & Nitz, D. A. (2022). Adaptive integration of self-motion and goals in posterior parietal cortex. *Cell Reports*, *38*(10), 110504. https://doi.org/10.1016/j.celrep.2022.110504

Amunts, K., Mohlberg, H., Bludau, S., & Zilles, K. (2020). Julich-Brain: A 3D probabilistic atlas of the human brain’s cytoarchitecture. Science, 369(6506), 988–992. https://doi.org/10.1126/science.abb4588

Andersen, R. A., & Buneo, C. A. (2002). Intentional Maps in Posterior Parietal Cortex. *Annual Review of Neuroscience*, *25*(1), 189–220. https://doi.org/10.1146/annurev.neuro.25.112701.142922

Andersson, Jesper L. R., Stefan Skare, and John Ashburner. 2003. “How to Correct Susceptibility Distortions in Spin-Echo Echo-Planar Images: Application to Diffusion Tensor Imaging.” NeuroImage 20 (2): 870–88. https://doi.org/10.1016/S1053-8119(03)00336-7.

Andersson, S. O., Moser, E. I., & Moser, M.-B. (2021). Visual stimulus features that elicit activity in object-vector cells. *Communications Biology*, *4*(1), Article 1. https://doi.org/10.1038/s42003-021-02727-5

Aronov, D., Nevers, R., & Tank, D. W. (2017). Mapping of a non-spatial dimension by the hippocampal/entorhinal circuit. *Nature*, *543*(7647), 719–722. https://doi.org/10.1038/nature21692

Avants, B. B., C. L. Epstein, M. Grossman, and J. C. Gee. 2008. “Symmetric Diffeomorphic Image Registration with Cross-Correlation: Evaluating Automated Labeling of Elderly and Neurodegenerative Brain.” Medical Image Analysis 12 (1): 26–41. https://doi.org/10.1016/j.media.2007.06.004.

Avila, E., Lakshminarasimhan, K. J., DeAngelis, G. C., & Angelaki, D. E. (2019). Visual and Vestibular Selectivity for Self-Motion in Macaque Posterior Parietal Area 7a. *Cerebral Cortex*, *29*(9), 3932–3947. https://doi.org/10.1093/cercor/bhy272

Bakermans, J. J. W., Warren, J., Whittington, J. C. R., & Behrens, T. E. J. (2025). Constructing future behavior in the hippocampal formation through composition and replay. *Nature Neuroscience*, *28*(5), 1061–1072. https://doi.org/10.1038/s41593-025-01908-3

Banino, A., Barry, C., Uria, B., Blundell, C., Lillicrap, T., Mirowski, P., Pritzel, A., Chadwick, M. J., Degris, T., Modayil, J., Wayne, G., Soyer, H., Viola, F., Zhang, B., Goroshin, R., Rabinowitz, N., Pascanu, R., Beattie, C., Petersen, S., … Kumaran, D. (2018). Vector-based navigation using grid-like representations in artificial agents. *Nature*, *557*(7705), 429–433. https://doi.org/10.1038/s41586-018-0102-6

Baram, A. B., Muller, T. H., Nili, H., Garvert, M. M., & Behrens, T. E. J. (2021). Entorhinal and ventromedial prefrontal cortices abstract and generalize the structure of reinforcement learning problems. *Neuron*, *109*(4), 713-723.e7. https://doi.org/10.1016/j.neuron.2020.11.024

Barnaveli, I., Viganò, S., Reznik, D., Haggard, P., & Doeller, C. F. (2025). Hippocampal-entorhinal cognitive maps and cortical motor system represent action plans and their outcomes. *Nature Communications*, *16*(1), 4139. https://doi.org/10.1038/s41467-025-59153-y

Barry, C., Hayman, R., Burgess, N., & Jeffery, K. J. (2007). Experience-dependent rescaling of entorhinal grids. *Nature Neuroscience*, *10*(6), Article 6. https://doi.org/10.1038/nn1905

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), 71–97. https://doi.org/10.1515/revneuro.2006.17.1-2.71

Behrens, T. E. J., Muller, T. H., Whittington, J. C. R., Mark, S., Baram, A. B., Stachenfeld, K. L., & Kurth-Nelson, Z. (2018). What Is a Cognitive Map? Organizing Knowledge for Flexible Behavior. *Neuron*, *100*(2), 490–509. https://doi.org/10.1016/j.neuron.2018.10.002

Behzadi, Yashar, Khaled Restom, Joy Liau, and Thomas T. Liu. 2007. "A Component Based Noise Correction Method (CompCor) for BOLD and Perfusion Based fMRI." NeuroImage 37 (1): 90–101. https://doi.org/10.1016/j.neuroimage.2007.04.042.

Bellmund, J. L. S., de Cothi, W., Ruiter, T. A., Nau, M., Barry, C., & Doeller, C. F. (2020). Deforming the metric of cognitive maps distorts memory. *Nature Human Behaviour*, *4*(2), Article 2. https://doi.org/10.1038/s41562-019-0767-3

Bellmund, J. L. S., Gärdenfors, P., Moser, E. I., & Doeller, C. F. (2018). Navigating cognition: Spatial codes for human thinking. *Science*, *362*(6415). https://doi.org/10.1126/science.aat6766

Bellmund, J. L., Deuker, L., & Doeller, C. F. (2019). Mapping sequence structure in the human lateral entorhinal cortex. eLife, 8, e45333. https://doi.org/10.7554/eLife.45333

Bellmund, J. L. S., Polti, I., & Doeller, C. F. (2020). Sequence Memory in the Hippocampal–Entorhinal Region. Journal of Cognitive Neuroscience, 32(11), 2056–2070. https://doi.org/10.1162/jocn_a_01592

Bernardi, S., Benna, M. K., Rigotti, M., Munuera, J., Fusi, S., & Salzman, C. D. (2020). The Geometry of Abstraction in the Hippocampus and Prefrontal Cortex. *Cell*. https://doi.org/10.1016/j.cell.2020.09.031

Bicanski, A., & Burgess, N. (2016). Environmental Anchoring of Head Direction in a Computational Model of Retrosplenial Cortex. *The Journal of Neuroscience*, *36*(46), 11601–11618. https://doi.org/10.1523/JNEUROSCI.0516-16.2016

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

Bicanski, A., & Burgess, N. (2019). A Computational Model of Visual Recognition Memory via Grid Cells. Current Biology, 29(6), 979-990.e4. https://doi.org/10.1016/j.cub.2019.01.077

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

Boccara, C. N., Nardin, M., Stella, F., O'Neill, J., & Csicsvari, J. (2019). The entorhinal cognitive map is attracted to goals. *Science*, *363*(6434), 1443–1447. https://doi.org/10.1126/science.aav4837

Boorman, E. D., Sweigart, S. C., & Park, S. A. (2021). Cognitive maps and novel inferences: A flexibility hierarchy. *Current Opinion in Behavioral Sciences*, *38*, 141–149. https://doi.org/10.1016/j.cobeha.2021.02.017

Bosch, J. J. van den, Golan, T., Peters, B., Taylor, J., Shahbazi, M., Lin, B., Charest, I., Diedrichsen, J., Kriegeskorte, N., Mur, M., & Schütt, H. H. (2025). A Python Toolbox for Representational Similarity Analysis. eLife, 14. https://doi.org/10.7554/eLife.107828.1

Bottini, R., & Doeller, C. F. (2020). Knowledge Across Reference Frames: Cognitive Maps and Image Spaces. *Trends in Cognitive Sciences*, *24*(8), 606–619. https://doi.org/10.1016/j.tics.2020.05.008

Braak, H., & Braak, E. (1991). Neuropathological stageing of Alzheimer-related changes. Acta Neuropathologica, 82(4), 239–259. https://doi.org/10.1007/BF00308809

Brun, V. H., Solstad, T., Kjelstrup, K. B., Fyhn, M., Witter, M. P., Moser, E. I., & Moser, M.-B. (2008). Progressive increase in grid scale from dorsal to ventral medial entorhinal cortex. *Hippocampus*, *18*(12), 1200–1212. https://doi.org/10.1002/hipo.20504

Buckmaster, C. A., Eichenbaum, H., Amaral, D. G., Suzuki, W. A., & Rapp, P. R. (2004). Entorhinal Cortex Lesions Disrupt the Relational Organization of Memory in Monkeys. The Journal of Neuroscience, 24(44), 9811–9825. https://doi.org/10.1523/JNEUROSCI.1532-04.2004

Burak, Y., & Fiete, I. (2006). Do We Understand the Emergent Dynamics of Grid Cell Activity? *Journal of Neuroscience*, *26*(37), 9352–9354. https://doi.org/10.1523/JNEUROSCI.2857-06.2006

Burak, Y., & Fiete, I. R. (2009). Accurate Path Integration in Continuous Attractor Network Models of Grid Cells. *PLoS Computational Biology*, *5*(2), e1000291. https://doi.org/10.1371/journal.pcbi.1000291

Burgess, N. (2006). Spatial memory: How egocentric and allocentric combine. *Trends in Cognitive Sciences*, *10*(12), 551–557. https://doi.org/10.1016/j.tics.2006.10.005

Bush, D., Barry, C., Manson, D., & Burgess, N. (2015). Using Grid Cells for Navigation. *Neuron*, *87*(3), 507–520. https://doi.org/10.1016/j.neuron.2015.07.006

Butler, W. N., Hardcastle, K., & Giocomo, L. M. (2019). Remembered reward locations restructure entorhinal spatial maps. *Science*, *363*(6434), 1447–1452. https://doi.org/10.1126/science.aav5297

Buzsáki, G., & Moser, E. I. (2013). Memory, navigation and theta rhythm in the hippocampal-entorhinal system. *Nature Neuroscience*, *16*(2), 130–138. https://doi.org/10.1038/nn.3304

Campbell, A. L., Naik, R. R., Sowards, L., & Stone, M. O. (2002). Biological infrared imaging and sensing. *Micron*, *33*(2), 211–225. https://doi.org/10.1016/S0968-4328(01)00010-5

Carpenter, F., Manson, D., Jeffery, K., Burgess, N., & Barry, C. (2015). Grid Cells Form a Global Representation of Connected Environments. *Current Biology*, *25*(9), 1176–1182. https://doi.org/10.1016/j.cub.2015.02.037

Chandra, S., Sharma, S., Chaudhuri, R., & Fiete, I. (2025). Episodic and associative memory from spatial scaffolds in the hippocampus. *Nature*, 1–13. https://doi.org/10.1038/s41586-024-08392-y

Chomsky, N. (1959). [Review of *Review of Verbal behavior*, by B. F. Skinner]. *Language*, *35*(1), 26–58. https://doi.org/10.2307/411334

Chun, M. M., & Turk-Browne, N. B. (2007). Interactions between attention and memory. *Current Opinion in Neurobiology*, *17*(2), 177–184. https://doi.org/10.1016/j.conb.2007.03.005

Cohen, N. J., & Eichenbaum, H. (1991). The theory that wouldn't die: A critical look at the spatial mapping theory of hippocampal function. Hippocampus, 1(3), 265–268. https://doi.org/10.1002/hipo.450010312

Constantinescu, A. O., O'Reilly, J. X., & Behrens, T. E. J. (2016). Organizing conceptual knowledge in humans with a gridlike code. *Science*, *352*(6292), 1464–1468. https://doi.org/10.1126/science.aaf0941

Courellis, H. S., Nummela, S. U., Metke, M., Diehl, G. W., Bussell, R., Cauwenberghs, G., & Miller, C. T. (2019). Spatial encoding in primate hippocampus during free navigation. *PLOS Biology*, *17*(12), e3000546. https://doi.org/10.1371/journal.pbio.3000546

Courellis, H. S., Minxha, J., Cardenas, A. R., Kimmel, D. L., Reed, C. M., Valiante, T. A., Salzman, C. D., Mamelak, A. N., Fusi, S., & Rutishauser, U. (2024). Abstract representations emerge in human

hippocampal neurons during inference. Nature, 632(8026), 841–849. https://doi.org/10.1038/s41586-024-07799-x

Cox, J., & Witten, I. B. (2019). Striatal circuits for reward learning and decision-making. *Nature Reviews Neuroscience*, *20*(8), 482–494. https://doi.org/10.1038/s41583-019-0189-2

Cox, Robert W., and James S. Hyde. 1997. "Software Tools for Analysis and Visualization of fMRI Data." NMR in Biomedicine 10 (4-5): 171–78. https://doi.org/10.1002/(SICI)1099-1492(199706/08)10:4/5<171::AID-NBM453>3.0.CO;2-L.

Crivelli-Decker, J., Clarke, A., Park, S. A., Huffman, D. J., Boorman, E. D., & Ranganath, C. (2023). Goal-oriented representations in the human hippocampus during planning and navigation. *Nature Communications*, *14*(1), 2946. https://doi.org/10.1038/s41467-023-35967-6

Crowe, D. A., Chafee, M. V., Averbeck, B. B., & Georgopoulos, A. P. (2004). Neural Activity in Primate Parietal Area 7a Related to Spatial Analysis of Visual Mazes. *Cerebral Cortex*, *14*(1), 23–34. https://doi.org/10.1093/cercor/bhg088

Cueva, C. J., & Wei, X.-X. (2018). *Emergence of grid-like representations by training recurrent neural networks to perform spatial localization* (No. arXiv:1803.07770). arXiv. https://doi.org/10.48550/arXiv.1803.07770

Danjo, T., Toyoizumi, T., & Fujisawa, S. (2018). Spatial representations of self and other in the hippocampus. *Science*, *359*(6372), 213–218. https://doi.org/10.1126/science.aao3898

Dayan, P. (2012). How to set the switches on this thing. Current Opinion in Neurobiology, 22(6), 1068–1074. https://doi.org/10.1016/j.conb.2012.05.011

Dehaene, S., Bossini, S., & Giraux, P. (1993). The mental representation of parity and number magnitude. Journal of Experimental Psychology: General, 122(3), 371–396. https://doi.org/10.1037/0096-3445.122.3.371

Derdikman, D., Whitlock, J. R., Tsao, A., Fyhn, M., Hafting, T., Moser, M.-B., & Moser, E. I. (2009). Fragmentation of grid cell maps in a multicompartment environment. *Nature Neuroscience*, *12*(10), 1325–1332. https://doi.org/10.1038/nn.2396

Deshmukh, S. S., & Knierim, J. J. (2013). Influence of local objects on hippocampal representations: Landmark vectors and memory. *Hippocampus*, *23*(4), 10.1002/hipo.22101. https://doi.org/10.1002/hipo.22101

Deubel, H., & Schneider, W. X. (1996). Saccade target selection and object recognition: Evidence for a common attentional mechanism. *Vision Research*, *36*(12), 1827–1837. https://doi.org/10.1016/0042-6989(95)00294-4

Diedrichsen, J., Provost, S., & Zareamoghaddam, H. (2016). On the distribution of cross-validated Mahalanobis distances (arXiv:1607.01371). arXiv. https://doi.org/10.48550/arXiv.1607.01371

Doeller, C. F., Barry, C., & Burgess, N. (2010). Evidence for grid cells in a human memory network. *Nature*, *463*(7281), 657–661. https://doi.org/10.1038/nature08704

Dong, L. L., & Fiete, I. R. (2024). Grid Cells in Cognition: Mechanisms and Function. *Annual Review of Neuroscience*. https://doi.org/10.1146/annurev-neuro-101323-112047

Dordek, Y., Soudry, D., Meir, R., & Derdikman, D. (2016). Extracting grid cell characteristics from place cell inputs using non-negative principal component analysis. *eLife*, *5*, e10094. https://doi.org/10.7554/eLife.10094

Dorrell, W., Latham, P. E., Behrens, T. E. J., & Whittington, J. C. R. (2022). *Actionable Neural Representations: Grid Cells from Minimal Constraints* (No. arXiv:2209.15563). arXiv. http://arxiv.org/abs/2209.15563

Eichenbaum, H. (2015). The Hippocampus as a Cognitive Map … of Social Space. *Neuron*, *87*(1), 9–11. https://doi.org/10.1016/j.neuron.2015.06.013

Eichenbaum, H., & Cohen, N. J. (2014). Can We Reconcile the Declarative Memory and Spatial Navigation Views on Hippocampal Function? *Neuron*, *83*(4), 764–770. https://doi.org/10.1016/j.neuron.2014.07.032

Eichenbaum, H., Dudchenko, P., Wood, E., Shapiro, M., & Tanila, H. (1999). The Hippocampus, Memory, and Place Cells: Is It Spatial Memory or a Memory Space? *Neuron*, *23*(2), 209–226. https://doi.org/10.1016/S0896-6273(00)80773-4

Ekstrom, A. D., Harootonian, S. K., & Huffman, D. J. (2020). Grid coding, spatial representation, and navigation: Should we assume an isomorphism? *Hippocampus*, *30*(4), 422–432. https://doi.org/10.1002/hipo.23175

Ekstrom, A. D., Kahana, M. J., Caplan, J. B., Fields, T. A., Isham, E. A., Newman, E. L., & Fried, I. (2003). Cellular networks underlying human spatial navigation. *Nature*, *425*(6954), 184–188. https://doi.org/10.1038/nature01964

Ekstrom, A. D., Spiers, H. J., Bohbot, V. D., & Rosenbaum, R. S. (2018). *Human spatial navigation*. Princeton University Press.

El-Gaby, M., Harris, A. L., Whittington, J. C. R., Dorrell, W., Bhomick, A., Walton, M. E., Akam, T., & Behrens, T. E. J. (2024). A cellular basis for mapping behavioural structure. *Nature*, *636*(8043), 671–680. https://doi.org/10.1038/s41586-024-08145-x

Eperon, A., Doeller, C. F., Theves, S., & Bottini, R. (2025). *Representation of conceptual affordances in eye movements and the entorhinal cortex* (p. 2025.01.21.633955). bioRxiv. https://doi.org/10.1101/2025.01.21.633955

Esteban, Oscar, Christopher Markiewicz, Ross W Blair, Craig Moodie, Ayse Ilkay Isik, Asier Erramuzpe Aliaga, James Kent, et al. 2018. “fMRIPrep: A Robust Preprocessing Pipeline for Functional MRI.” Nature Methods. https://doi.org/10.1038/s41592-018-0235-4.

Esteban, Oscar, Ross Blair, Christopher J. Markiewicz, Shoshana L. Berleant, Craig Moodie, Feilong Ma, Ayse Ilkay Isik, et al. 2018. “fMRIPrep 22.0.1.” Software. https://doi.org/10.5281/zenodo.852659.

Etienne, A. S., & Jeffery, K. J. (2004). Path integration in mammals. *Hippocampus*, *14*(2), 180–192. https://doi.org/10.1002/hipo.10173

Evans, AC, AL Janke, DL Collins, and S Baillet. 2012. “Brain Templates and Atlases.” NeuroImage 62 (2): 911–22. https://doi.org/10.1016/j.neuroimage.2012.01.024.

Fernandez-Velasco, P., & Spiers, H. J. (2024). Wayfinding across ocean and tundra: What traditional cultures teach us about navigation. Trends in Cognitive Sciences, 28(1), 56–71. https://doi.org/10.1016/j.tics.2023.09.004

Fodor, J. A. (1975). *The Language of Thought*. Harvard University Press.

Fonov, VS, AC Evans, RC McKinstry, CR Almli, and DL Collins. 2009. “Unbiased Nonlinear Average Age-Appropriate Brain Templates from Birth to Adulthood.” NeuroImage 47, Supplement 1: S102. https://doi.org/10.1016/S1053-8119(09)70884-5.

Frank, L. M., Brown, E. N., & Wilson, M. (2000). Trajectory Encoding in the Hippocampus and Entorhinal Cortex. *Neuron*, *27*(1), 169–178. https://doi.org/10.1016/S0896-6273(00)00018-0

Frankland, S. M., & Greene, J. D. (2020). Concepts and Compositionality: In Search of the Brain’s Language of Thought. *Annual Review of Psychology*, *71*, 273–303. https://doi.org/10.1146/annurev-psych-122216-011829

Friston, K., & Kiebel, S. (2009). Predictive coding under the free-energy principle. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1521), 1211–1221. https://doi.org/10.1098/rstb.2008.0300

Fuhs, M. C., & Touretzky, D. S. (2006). A Spin Glass Model of Path Integration in Rat Medial Entorhinal Cortex. *The Journal of Neuroscience*, *26*(16), 4266–4276. https://doi.org/10.1523/jneurosci.4353-05.2006

Fyhn, M., Hafting, T., Treves, A., Moser, M.-B., & Moser, E. I. (2007). Hippocampal remapping and grid realignment in entorhinal cortex. *Nature*, *446*(7132), 190–194. https://doi.org/10.1038/nature05601

Garcia, A. D., & Buffalo, E. A. (2020). Anatomy and Function of the Primate Entorhinal Cortex. Annual Review of Vision Science, 6(Volume 6, 2020), 411–432. https://doi.org/10.1146/annurev-vision-030320-041115

Gärdenfors, P. (2000). *Conceptual spaces: The geometry of thought*. MIT Press.

Gardner, R. J., Hermansen, E., Pachitariu, M., Burak, Y., Baas, N. A., Dunn, B. A., Moser, M.-B., & Moser, E. I. (2022). Toroidal topology of population activity in grid cells. *Nature*, 1–6. https://doi.org/10.1038/s41586-021-04268-7

Garvert, M. M., Dolan, R. J., & Behrens, T. E. (2017). A map of abstract relational knowledge in the human hippocampal–entorhinal cortex. *eLife*, *6*, e17086. https://doi.org/10.7554/eLife.17086

Garvert, M. M., Saanum, T., Schulz, E., Schuck, N. W., & Doeller, C. F. (2023). Hippocampal spatio-predictive cognitive maps adaptively guide reward generalization. Nature Neuroscience, 26(4), Article 4. https://doi.org/10.1038/s41593-023-01283-x

George, D., Rikhye, R. V., Gothoskar, N., Guntupalli, J. S., Dedieu, A., & Lázaro-Gredilla, M. (2021). Clone-structured graph representations enable flexible learning and vicarious evaluation of cognitive maps. *Nature Communications*, *12*(1), Article 1. https://doi.org/10.1038/s41467-021-22559-5

Gershman, S. J., & Daw, N. D. (2017). Reinforcement Learning and Episodic Memory in Humans and Animals: An Integrative Framework. Annual Review of Psychology, 68(Volume 68, 2017), 101–128. https://doi.org/10.1146/annurev-psych-122414-033625

Gershman, S. J. (2018). The Successor Representation: Its Computational Logic and Neural Substrates. *Journal of Neuroscience*, *38*(33), 7193–7200. https://doi.org/10.1523/JNEUROSCI.0151-18.2018

Giari, G., Vignali, L., Xu, Y., & Bottini, R. (2023). MEG frequency tagging reveals a grid-like code during attentional movements. *Cell Reports*, *42*(10), 113209. https://doi.org/10.1016/j.celrep.2023.113209

Gibson, J. J. (1979). The ecological approach to visual perception (pp. xiv, 332). Houghton, Mifflin and Company.

Gibson, J. J. (1979). *The ecological approach to visual perception* (pp. xiv, 332). Houghton, Mifflin and Company.

Ginosar, G., Aljadeff, J., Burak, Y., Sompolinsky, H., Las, L., & Ulanovsky, N. (2021). Locally ordered representation of 3D space in the entorhinal cortex. *Nature*, *596*(7872), 404–409. https://doi.org/10.1038/s41586-021-03783-x

Ginosar, G., Aljadeff, J., Las, L., Derdikman, D., & Ulanovsky, N. (2023). Are grid cells used for navigation? On local metrics, subjective spaces, and black holes. *Neuron*, *111*(12), 1858–1875. https://doi.org/10.1016/j.neuron.2023.03.027

Giocomo, L. M., Stensola, T., Bonnevie, T., Van Cauter, T., Moser, M.-B., & Moser, E. I. (2014). Topography of Head Direction Cells in Medial Entorhinal Cortex. *Current Biology*, *24*(3), 252–262. https://doi.org/10.1016/j.cub.2013.12.002

Gonzalez, A., & Giocomo, L. M. (2024). Parahippocampal neurons encode task-relevant information for goal-directed navigation. *eLife*, *12*, RP85646. https://doi.org/10.7554/eLife.85646

Gorgolewski, K., C. D. Burns, C. Madison, D. Clark, Y. O. Halchenko, M. L. Waskom, and S. Ghosh. 2011. "Nipype: A Flexible, Lightweight and Extensible Neuroimaging Data Processing Framework in Python." Frontiers in Neuroinformatics 5: 13. https://doi.org/10.3389/fninf.2011.00013.

Gorgolewski, Krzysztof J., Oscar Esteban, Christopher J. Markiewicz, Erik Ziegler, David Gage Ellis, Michael Philipp Notter, Dorota Jarecka, et al. 2018. "Nipype." Software. https://doi.org/10.5281/zenodo.596855.

Graichen, L. P., Linder, M. S., Keuter, L., Jensen, O., Doeller, C. F., Lamm, C., Staudigl, T., & Wagner, I. C. (2024). *Entorhinal grid-like codes for visual space during memory formation* (p. 2024.09.27.615339). bioRxiv. https://doi.org/10.1101/2024.09.27.615339

Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., Goj, R., Jas, M., Brooks, T., Parkkonen, L., & Hämäläinen, M. (2013). MEG and EEG data analysis with MNE-Python. Frontiers in Neuroscience, 7. https://doi.org/10.3389/fnins.2013.00267

Greve, Douglas N, and Bruce Fischl. 2009. "Accurate and Robust Brain Image Alignment Using Boundary-Based Registration." NeuroImage 48 (1): 63–72. https://doi.org/10.1016/j.neuroimage.2009.06.060.

Grieves, R. M., Jedidi-Ayoub, S., Mishchanchuk, K., Liu, A., Renaudineau, S., Duvelle, É., & Jeffery, K. J. (2021). Irregular distribution of grid cell firing fields in rats exploring a 3D volumetric space. *Nature Neuroscience*, *24*(11), 1567–1573. https://doi.org/10.1038/s41593-021-00907-4

Groner, R., & Groner, M. T. (1989). Attention and eye movement control: An overview. *European Archives of Psychiatry and Neurological Sciences*, *239*(1), 9–16. https://doi.org/10.1007/BF01739737

Guanella, A., Kiper, D., & Verschure, P. (2007). A model of grid cells based on a twisted torus topology. *International Journal of Neural Systems*, *17*(04), 231–240. https://doi.org/10.1142/s0129065707001093

Hafting, T., Fyhn, M., Molden, S., Moser, M.-B., & Moser, E. I. (2005). Microstructure of a spatial map in the entorhinal cortex. *Nature*, *436*(7052), Article 7052. https://doi.org/10.1038/nature03721

Hales, J. B., Schlesiger, M. I., Leutgeb, J. K., Squire, L. R., Leutgeb, S., & Clark, R. E. (2014). Medial Entorhinal Cortex Lesions Only Partially Disrupt Hippocampal Place Cells and Hippocampus-Dependent Place Memory. *Cell Reports*, *9*(3), 893–901. https://doi.org/10.1016/j.celrep.2014.10.009

Hartmann, M. (2022). Summing up: A functional role of eye movements along the mental number line for arithmetic. *Acta Psychologica*, *230*, 103770. https://doi.org/10.1016/j.actpsy.2022.103770

Hassabis, D., & Maguire, E. A. (2009). The construction system of the brain. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1521), 1263–1271. https://doi.org/10.1098/rstb.2008.0296

Hassabis, D., Kumaran, D., Vann, S. D., & Maguire, E. A. (2007). Patients with hippocampal amnesia cannot imagine new experiences. *Proceedings of the National Academy of Sciences*, *104*(5), 1726–1731. https://doi.org/10.1073/pnas.0610561104

He, Q., & Brown, T. I. (2019). Environmental Barriers Disrupt Grid-like Representations in Humans during Navigation. *Current Biology*, *29*(16), 2718-2722.e3. https://doi.org/10.1016/j.cub.2019.06.072

Heilbron, M., Armeni, K., Schoffelen, J.-M., Hagoort, P., & De Lange, F. P. (2022). A hierarchy of linguistic predictions during natural language comprehension. Proceedings of the National Academy of Sciences, 119(32), e2201968119. https://doi.org/10.1073/pnas.2201968119

Hindy, N. C., & Turk-Browne, N. B. (2016). Action-Based Learning of Multistate Objects in the Medial Temporal Lobe. Cerebral Cortex (New York, NY), 26(5), 1853–1865. https://doi.org/10.1093/cercor/bhv030

Hoffman, J. E., & Subramaniam, B. (1995). The role of visual attention in saccadic eye movements. *Perception & Psychophysics*, *57*(6), 787–795. https://doi.org/10.3758/BF03206794

Horner, A. J., Bisby, J. A., Zotow, E., Bush, D., & Burgess, N. (2016). Grid-like Processing of Imagined Navigation. Current Biology, 26(6), 842–847. https://doi.org/10.1016/j.cub.2016.01.042

Howard, L. R., Javadi, A. H., Yu, Y., Mill, R. D., Morrison, L. C., Knight, R., Loftus, M. M., Staskute, L., & Spiers, H. J. (2014). The Hippocampus and Entorhinal Cortex Encode the Path and Euclidean Distances to Goals during Navigation. Current Biology, 24(12), 1331–1340. https://doi.org/10.1016/j.cub.2014.05.001

Høydal, Ø. A., Skytøen, E. R., Andersson, S. O., Moser, M.-B., & Moser, E. I. (2019). Object-vector coding in the medial entorhinal cortex. *Nature*, *568*(7752), Article 7752. https://doi.org/10.1038/s41586-019-1077-7

Hwang, J., Neupane, S., Jazayeri, M., & Fiete, I. (2023). A Grid Cell-Place Cell Scaffold Allows Rapid Learning and Generalization at Multiple Levels on Mental Navigation Tasks. 2023 Conference on Cognitive Computational Neuroscience. 2023 Conference on Cognitive Computational Neuroscience, Oxford, UK. https://doi.org/10.32470/CCN.2023.1681-0

Iyer, A., Chandra, S., Sharma, S., & Fiete, I. (2024). Flexible mapping of abstract domains by grid cells via self-supervised extraction and projection of generalized velocity signals. *Advances in Neural Information Processing Systems*, *37*, 85441–85466.

Jeffery, K. J. (2021). How environmental movement constraints shape the neural code for space. *Cognitive Processing*, *22*(Suppl 1), 97–104. https://doi.org/10.1007/s10339-021-01045-2

Jenkinson, Mark, and Stephen Smith. 2001. “A Global Optimisation Method for Robust Affine Registration of Brain Images.” Medical Image Analysis 5 (2): 143–56. https://doi.org/10.1016/S1361-8415(01)00036-6.

Jenkinson, Mark, Peter Bannister, Michael Brady, and Stephen Smith. 2002. “Improved Optimization for the Robust and Accurate Linear Registration and Motion Correction of Brain Images.” NeuroImage 17 (2): 825–41. https://doi.org/10.1006/nimg.2002.1132.

Jones, E. A. A., Low, I. I. C., Cho, F. S., & Giocomo, L. M. (2024). *Entorhinal cortex represents task-relevant remote locations independent of CA1* (p. 2024.07.23.604815). bioRxiv. https://doi.org/10.1101/2024.07.23.604815

Julian, J. B., Keinath, A. T., Frazzetta, G., & Epstein, R. A. (2018). Human entorhinal cortex represents visual space using a boundary-anchored grid. *Nature Neuroscience*, *21*(2), Article 2. https://doi.org/10.1038/s41593-017-0049-1

Kaiser, D., Jacobs, A. M., & Cichy, R. M. (2022). Modelling brain representations of abstract concepts. PLOS Computational Biology, 18(2), e1009837. https://doi.org/10.1371/journal.pcbi.1009837

Kerrén, C., Michelmann, S., & Doeller, C. F. (2025). *Hippocampal ripples initiate cortical dimensionality expansion for memory retrieval* (p. 2025.04.22.649929). bioRxiv. https://doi.org/10.1101/2025.04.22.649929

Khona, M., & Fiete, I. R. (2022). Attractor and integrator networks in the brain. *Nature Reviews Neuroscience*, *23*(12), Article 12. https://doi.org/10.1038/s41583-022-00642-0

Killian, N. J., & Buffalo, E. A. (2018). Grid cells map the visual world. *Nature Neuroscience*, *21*(2), Article 2. https://doi.org/10.1038/s41593-017-0062-4

Killian, N. J., Jutras, M. J., & Buffalo, E. A. (2012). A map of visual space in the primate entorhinal cortex. *Nature*, *491*(7426), Article 7426. https://doi.org/10.1038/nature11587

Killian, N. J., Potter, S. M., & Buffalo, E. A. (2015). Saccade direction encoding in the primate entorhinal cortex during visual exploration. *Proceedings of the National Academy of Sciences*, *112*(51), 15743–15748. https://doi.org/10.1073/pnas.1417059112

Klingberg, T., Roland, P. E., & Kawashima, R. (1994). The human entorhinal cortex participates in associative memory. Neuroreport, 6(1), 57–60. https://doi.org/10.1097/00001756-199412300-00016

Knowlton, B. J., & Squire, L. R. (1993). The Learning of Categories: Parallel Brain Systems for Item Memory and Category Knowledge. Science, 262(5140), 1747–1749. https://doi.org/10.1126/science.8259522

Knudsen, E. B., & Wallis, J. D. (2021). Hippocampal neurons construct a map of an abstract value space. *Cell*, *184*(18), 4640-4650.e10. https://doi.org/10.1016/j.cell.2021.07.010

Kolibius, L. D., Roux, F., Parish, G., Ter Wal, M., Van Der Plas, M., Chelvarajah, R., Sawlani, V., Rollings, D. T., Lang, J. D., Gollwitzer, S., Walther, K., Hopfengärtner, R., Kreiselmeyer, G., Hamer, H., Staresina, B. P., Wimber, M., Bowman, H., & Hanslmayr, S. (2023). Hippocampal neurons code individual episodic memories in humans. Nature Human Behaviour, 7(11), 1968–1979. https://doi.org/10.1038/s41562-023-01706-6

Kriegeskorte, N., Goebel, R., & Bandettini, P. (2006). Information-based functional brain mapping. Proceedings of the National Academy of Sciences, 103(10), 3863–3868. https://doi.org/10.1073/pnas.0600244103

Kriegeskorte, N., Mur, M., & Bandettini, P. (2008). Representational similarity analysis—Connecting the branches of systems neuroscience. Frontiers in Systems Neuroscience, 2. https://www.frontiersin.org/articles/10.3389/neuro.06.004.2008

Kropff, E., Carmichael, J. E., Moser, M.-B., & Moser, E. I. (2015). Speed cells in the medial entorhinal cortex. Nature, 523(7561), Article 7561. https://doi.org/10.1038/nature14622

Krupic, J., Bauza, M., Burton, S., & O'Keefe, J. (2016). Framing the grid: Effect of boundaries on grid cells and navigation. *The Journal of Physiology*, *594*(22), 6489–6499. https://doi.org/10.1113/JP270607

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

Ku, S., Hargreaves, E. L., Wirth, S., & Suzuki, W. A. (2021). The contributions of entorhinal cortex and hippocampus to error driven learning. Communications Biology, 4(1), 618. https://doi.org/10.1038/s42003-021-02096-z

Kutter, E. F., Boström, J., Elger, C. E., Nieder, A., & Mormann, F. (2022). Neuronal codes for arithmetic rule processing in the human brain. Current Biology, 32(6), 1275-1284.e4. https://doi.org/10.1016/j.cub.2022.01.054

Lanczos, C. 1964. "Evaluation of Noisy Data." Journal of the Society for Industrial and Applied Mathematics Series B Numerical Analysis 1 (1): 76–85. https://doi.org/10.1137/0701007.

Liang, Z., Wu, S., Wu, J., Wang, W.-X., Qin, S., & Liu, C. (2024). Distance and grid-like codes support the navigation of abstract social space in the human brain. *eLife*, *12*, RP89025. https://doi.org/10.7554/eLife.89025

Lindsay, G. (2021). *Models of the Mind: How Physics, Engineering and Mathematics Have Shaped Our Understanding of the Brain*. Bloomsbury Publishing.

Loetscher, T., Bockisch, C. J., & Brugger, P. (2008). Looking for the answer: The mind's eye in number space. Neuroscience, 151(3), 725–729. https://doi.org/10.1016/j.neuroscience.2007.07.068

Loetscher, T., Bockisch, C. J., Nicholls, M. E. R., & Brugger, P. (2010). Eye position predicts what number you have in mind. *Current Biology*, *20*(6), R264–R265. https://doi.org/10.1016/j.cub.2010.01.015

Lohmann, K. J. (2010). Magnetic-field perception. *Nature*, *464*(7292), 1140–1142. https://doi.org/10.1038/4641140a

Lombardi, C. M., & Hurlbert, S. H. (2009). Misprescription and misuse of one-tailed tests. Austral Ecology, 34(4), 447–468. https://doi.org/10.1111/j.1442-9993.2009.01946.x

Luo, X., Mok, R. M., & Love, B. C. (2024). The inevitability and superfluousness of cell types in spatial cognition. *eLife*, *13*. https://doi.org/10.7554/eLife.99047.1

Maguire, E. A., Nannery, R., & Spiers, H. J. (2006). Navigation around London by a taxi driver with bilateral hippocampal lesions. Brain, 129(11), 2894–2907. https://doi.org/10.1093/brain/awl286

Mallory, C. S., Hardcastle, K., Campbell, M. G., Attinger, A., Low, I. I. C., Raymond, J. L., & Giocomo, L. M. (2021). Mouse entorhinal cortex encodes a diverse repertoire of self-motion signals. *Nature Communications*, *12*(1), Article 1. https://doi.org/10.1038/s41467-021-20936-8

Manns, J. R., & Eichenbaum, H. (2006). Evolution of declarative memory. Hippocampus, 16(9), 795–808. https://doi.org/10.1002/hipo.20205

Mark, S., Schwartenbeck, P., Hahamy, A., Samborska, V., Baram, A. B., & Behrens, T. E. (2024). Flexible neural representations of abstract structural knowledge in the human Entorhinal Cortex. *eLife*, *13*. https://doi.org/10.7554/eLife.101134.1

Mathis, A., Stemmler, M. B., & Herz, A. V. (2015). Probable nature of higher-dimensional symmetries underlying mammalian grid-cell activity patterns. *eLife*, *4*, e05979. https://doi.org/10.7554/eLife.05979

McAvan, A. S., Wank, A. A., Rapcsak, S. Z., Grilli, M. D., & Ekstrom, A. D. (2022). Largely intact memory for spatial locations during navigation in an individual with dense amnesia. Neuropsychologia, 170, 108225. https://doi.org/10.1016/j.neuropsychologia.2022.108225

McDougle, S. D., Wilterson, S. A., Turk-Browne, N. B., & Taylor, J. A. (2022). Revisiting the Role of the Medial Temporal Lobe in Motor Learning. Journal of Cognitive Neuroscience, 34(3), 532–549. https://doi.org/10.1162/jocn_a_01809

McNaughton, B. L., Battaglia, F. P., Jensen, O., Moser, E. I., & Moser, M.-B. (2006). Path integration and the neural basis of the "cognitive map." *Nature Reviews Neuroscience*, *7*(8), Article 8. https://doi.org/10.1038/nrn1932

Mehta, M. R., Barnes, C. A., & McNaughton, B. L. (1997). Experience-dependent, asymmetric expansion of hippocampal place fields. *Proceedings of the National Academy of Sciences*, *94*(16), 8918–8921. https://doi.org/10.1073/pnas.94.16.8918

Mehta, M. R., Quirk, M. C., & Wilson, M. A. (2000). Experience-Dependent Asymmetric Shape of Hippocampal Receptive Fields. *Neuron*, *25*(3), 707–715. https://doi.org/10.1016/S0896-6273(00)81072-7

Meister, M. L. R., & Buffalo, E. A. (2016). Getting directions from the hippocampus: The neural connection between looking and memory. *Neurobiology of Learning and Memory*, *134*, 135–144. https://doi.org/10.1016/j.nlm.2015.12.004

Miller, E. K., & Cohen, J. D. (2001). An Integrative Theory of Prefrontal Cortex Function. *Annual Review of Neuroscience*, *24*(1), 167–202. https://doi.org/10.1146/annurev.neuro.24.1.167

Momennejad, I., Russek, E. M., Cheong, J. H., Botvinick, M. M., Daw, N. D., & Gershman, S. J. (2017). The successor representation in human reinforcement learning. *Nature Human Behaviour*, *1*(9), 680–692. https://doi.org/10.1038/s41562-017-0180-8

Moss, C. F., Ortiz, S. T., & Wahlberg, M. (2023). Adaptive echolocation behavior of bats and toothed whales in dynamic soundscapes. *Journal of Experimental Biology*, *226*(9), jeb245450. https://doi.org/10.1242/jeb.245450

Muller, R., & Kubie, J. (1987). The effects of changes in the environment on the spatial firing of hippocampal complex-spike cells. *The Journal of Neuroscience*, *7*(7), 1951–1968. https://doi.org/10.1523/JNEUROSCI.07-07-01951.1987

Munn, R. G. K., Mallory, C. S., Hardcastle, K., Chetkovich, D. M., & Giocomo, L. M. (2020). Entorhinal velocity signals reflect environmental geometry. *Nature Neuroscience*, *23*(2), 239–251. https://doi.org/10.1038/s41593-019-0562-5

Muzzio, I. A., Kentros, C., & Kandel, E. (2009). What is remembered? Role of attention on the encoding and retrieval of hippocampal representations. *The Journal of Physiology*, *587*(12), 2837–2854. https://doi.org/10.1113/jphysiol.2009.172445

Nau, M. (2022). Functional imaging of the human medial temporal lobe. https://doi.org/10.17605/OSF.IO/CQN4Z

Nau, M., Julian, J. B., & Doeller, C. F. (2018). How the Brain's Navigation System Shapes Our Visual Experience. *Trends in Cognitive Sciences*, *22*(9), 810–825. https://doi.org/10.1016/j.tics.2018.06.008

Nau, M., Navarro Schröder, T., Bellmund, J. L. S., & Doeller, C. F. (2018). Hexadirectional coding of visual space in human entorhinal cortex. *Nature Neuroscience*, *21*(2), Article 2. https://doi.org/10.1038/s41593-017-0050-8

Neupane, S., Fiete, I., & Jazayeri, M. (2024). Mental navigation in the primate entorhinal cortex. *Nature*, *630*(8017), 704–711. https://doi.org/10.1038/s41586-024-07557-z

Newton, C., Pope, M., Rua, C., Henson, R., Ji, Z., Burgess, N., Rodgers, C. T., Stangl, M., Dounavi, M.-E., Castegnaro, A., Koychev, I., Malhotra, P., Wolbers, T., Ritchie, K., Ritchie, C. W., O'Brien, J., Su, L., Chan, D., & Programme, for the P. D. R. (2024). Entorhinal-based path integration selectively predicts midlife risk of Alzheimer's disease. Alzheimer's & Dementia, 20(4), 2779–2793. https://doi.org/10.1002/alz.13733

Nili, H., Wingfield, C., Walther, A., Su, L., Marslen-Wilson, W., & Kriegeskorte, N. (2014). A Toolbox for Representational Similarity Analysis. PLOS Computational Biology, 10(4), e1003553. https://doi.org/10.1371/journal.pcbi.1003553

Nitsch, A., Garvert, M. M., Bellmund, J. L. S., Schuck, N. W., & Doeller, C. F. (2024). Grid-like entorhinal representation of an abstract value space during prospective decision making. *Nature Communications*, *15*(1), 1198. https://doi.org/10.1038/s41467-024-45127-z

Nitz, D. A. (2014). The Posterior Parietal Cortex: Interface Between Maps of External Spaces and the Generation of Action Sequences. In D. Derdikman & J. J. Knierim (Eds.), *Space,Time and Memory in the Hippocampal Formation* (pp. 27–54). Springer. https://doi.org/10.1007/978-3-7091-1292-2_2

Nour, M. M., Liu, Y., Arumuham, A., Kurth-Nelson, Z., & Dolan, R. J. (2021). Impaired neural replay of inferred relationships in schizophrenia. *Cell*, *184*(16), 4315-4328.e17. https://doi.org/10.1016/j.cell.2021.06.012

Núñez, R., & Fias, W. (2017). Ancestral Mental Number Lines: What Is the Evidence? Cognitive Science, 41(8), 2262–2266. https://doi.org/10.1111/cogs.12296

O'Brien, K., Krüger, G., Lazeyras, F., Gruetter, R., & Roche, A. (2013). A simple method to denoise MP2RAGE. https://infoscience.epfl.ch/handle/20.500.14299/92057

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

O'Keefe, J., & Dostrovsky, J. (1971). The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. *Brain Research*, *34*(1), 171–175. https://doi.org/10.1016/0006-8993(71)90358-1

O'Keefe, J., & Nadel, L. (1978). *The hippocampus as a cognitive map*. Clarendon Press ; Oxford University Press.

Ólafsdóttir, H. F., Bush, D., & Barry, C. (2018). The Role of Hippocampal Replay in Memory and Planning. *Current Biology*, *28*(1), R37–R50. https://doi.org/10.1016/j.cub.2017.10.073

Omer, D. B., Maimon, S. R., Las, L., & Ulanovsky, N. (2018). Social place-cells in the bat hippocampus. *Science*, *359*(6372), 218–224. https://doi.org/10.1126/science.aao3474

Ormond, J., & O'Keefe, J. (2022). Hippocampal place cells have goal-oriented vector fields during navigation. *Nature*, *607*(7920), 741–746. https://doi.org/10.1038/s41586-022-04913-9

Park, S. A., Miller, D. S., & Boorman, E. D. (2021). Inferences on a multidimensional social hierarchy use a grid-like code. *Nature Neuroscience*, *24*(9), Article 9. https://doi.org/10.1038/s41593-021-00916-3

Park, S. A., Miller, D. S., Nili, H., Ranganath, C., & Boorman, E. D. (2020). Map Making: Constructing, Combining, and Inferring on Abstract Cognitive Maps. *Neuron*, *107*(6), 1226-1238.e8. https://doi.org/10.1016/j.neuron.2020.06.030

Patriat, Rémi, Richard C. Reynolds, and Rasmus M. Birn. 2017. "An Improved Model of Motion-Related Signal Changes in fMRI." NeuroImage 144, Part A (January): 74–82. https://doi.org/10.1016/j.neuroimage.2016.08.051.

Payne, H. L., Lynch, G. F., & Aronov, D. (2021). Neural representations of space in the hippocampus of a food-caching bird. *Science*, *373*(6552), 343–348. https://doi.org/10.1126/science.abg2009

Peer, M., Brunec, I. K., Newcombe, N. S., & Epstein, R. A. (2021). Structuring Knowledge with Cognitive Maps and Cognitive Graphs. Trends in Cognitive Sciences, 25(1), 37–54. https://doi.org/10.1016/j.tics.2020.10.004

Perentos, N., Krstulovic, M., & Morton, A. J. (2022). Deep brain electrophysiology in freely moving sheep. *Current Biology*, *32*(4), 763-774.e4. https://doi.org/10.1016/j.cub.2021.12.035

Piazza, M., & Izard, V. (2009). How Humans Count: Numerosity and the Parietal Cortex. The Neuroscientist, 15(3), 261–273. https://doi.org/10.1177/1073858409333073

Possidónio, C., Graça, J., Piazza, J., & Prada, M. (2019). Animal Images Database: Validation of 120 Images for Human-Animal Studies. *Animals*, *9*(8), Article 8. https://doi.org/10.3390/ani9080475

Power, Jonathan D., Anish Mitra, Timothy O. Laumann, Abraham Z. Snyder, Bradley L. Schlaggar, and Steven E. Petersen. 2014. "Methods to Detect, Characterize, and Remove Motion Artifact in Resting State fMRI." NeuroImage 84 (Supplement C): 320–41. https://doi.org/10.1016/j.neuroimage.2013.08.048.

Pruim, Raimon H. R., Maarten Mennes, Daan van Rooij, Alberto Llera, Jan K. Buitelaar, and Christian F. Beckmann. 2015. "ICA-AROMA: A Robust ICA-Based Strategy for Removing Motion Artifacts from fMRI Data." NeuroImage 112 (Supplement C): 267–77. https://doi.org/10.1016/j.neuroimage.2015.02.064.

Pudhiyidath, A., Morton, N. W., Duran, R. V., Schapiro, A. C., Momennejad, I., Hinojosa-Rowland, D. M., Molitor, R. J., & Preston, A. R. (2021). Representations of temporal community structure in hippocampus and precuneus predict inductive reasoning decisions (p. 2021.10.12.462707). https://doi.org/10.1101/2021.10.12.462707

Qasim, S. E., Fried, I., & Jacobs, J. (2021). Phase precession in the human hippocampus and entorhinal cortex. *Cell*, *184*(12), 3242-3255.e10. https://doi.org/10.1016/j.cell.2021.04.017

Qasim, S. E., Reinacher, P. C., Brandt, A., Schulze-Bonhage, A., & Kunz, L. (2023). *Neurons in the human entorhinal cortex map abstract emotion space* (p. 2023.08.10.552884). bioRxiv. https://doi.org/10.1101/2023.08.10.552884

Rao, R. P. N., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, *2*(1), 79–87. https://doi.org/10.1038/4580

Redish, A. D., Elga, A. N., & Touretzky, D. S. (1996). A coupled attractor model of the rodent head direction system. *Network: Computation in Neural Systems*, *7*(4), 671–685. https://doi.org/10.1088/0954-898X_7_4_004

Restle, F. (1970). Speed of adding and comparing numbers. Journal of Experimental Psychology, 83(2, Pt.1), 274–278. https://doi.org/10.1037/h0028573

Ringo, J. L., Sobotka, S., Diltz, M. D., & Bunce, C. M. (1994). Eye movements modulate activity in hippocampal, parahippocampal, and inferotemporal neurons. Journal of Neurophysiology, 71(3), 1285–1288. https://doi.org/10.1152/jn.1994.71.3.1285

Sakurai, Y. (2002). Coding of auditory temporal and pitch information by hippocampal individual cells and cell assemblies in the rat. *Neuroscience*, *115*(4), 1153–1163. https://doi.org/10.1016/S0306-4522(02)00509-2

Samsonovich, A., & McNaughton, B. L. (1997). Path Integration and Cognitive Mapping in a Continuous Attractor Neural Network Model. *Journal of Neuroscience*, *17*(15), 5900–5920. https://doi.org/10.1523/JNEUROSCI.17-15-05900.1997

Sarel, A., Finkelstein, A., Las, L., & Ulanovsky, N. (2017). Vectorial representation of spatial goals in the hippocampus of bats. *Science*, *355*(6321), 176–180. https://doi.org/10.1126/science.aak9589

Sargolini, F., Fyhn, M., Hafting, T., McNaughton, B. L., Witter, M. P., Moser, M.-B., & Moser, E. I. (2006). Conjunctive Representation of Position, Direction, and Velocity in Entorhinal Cortex. *Science*, *312*(5774), 758–762. https://doi.org/10.1126/science.1125572

Satterthwaite, Theodore D., Mark A. Elliott, Raphael T. Gerraty, Kosha Ruparel, James Loughead, Monica E. Calkins, Simon B. Eickhoff, et al. 2013. "An improved framework for confound regression and filtering for control of motion artifact in the preprocessing of resting-state functional connectivity data." NeuroImage 64 (1): 240–56. https://doi.org/10.1016/j.neuroimage.2012.08.052.

Save, E., & Sargolini, F. (2017). Disentangling the Role of the MEC and LEC in the Processing of Spatial and Non-Spatial Information: Contribution of Lesion Studies. *Frontiers in Systems Neuroscience*, *11*, 81. https://doi.org/10.3389/fnsys.2017.00081

Schaeffer, R., Khona, M., & Fiete, I. R. (2022). *No Free Lunch from Deep Learning in Neuroscience: A Case Study through Models of the Entorhinal-Hippocampal Circuit* (p. 2022.08.07.503109). bioRxiv. https://doi.org/10.1101/2022.08.07.503109

Schafer, M., & Schiller, D. (2018). Navigating Social Space. *Neuron*, *100*(2), 476–489. https://doi.org/10.1016/j.neuron.2018.10.006

Schafer, M., Kamilar-Britt, P., Sahani, V., Bachi, K., & Schiller, D. (2022). Hippocampal Place-like Signal in Latent Space (p. 2022.07.15.499827). bioRxiv. https://doi.org/10.1101/2022.07.15.499827

Schäfer, T. A. J., Thalmann, M., Schulz, E., Doeller, C. F., & Theves, S. (2024). The hippocampus supports interpolation into new states during category abstraction (p. 2024.05.14.594185). bioRxiv. https://doi.org/10.1101/2024.05.14.594185

Schapiro, A. C., Kustner, L. V., & Turk-Browne, N. B. (2012). Shaping of Object Representations in the Human Medial Temporal Lobe Based on Temporal Regularities. Current Biology, 22(17), 1622–1627. https://doi.org/10.1016/j.cub.2012.06.056

Shaki, S., Fischer, M. H., & Petrusic, W. M. (2009). Reading habits for both words and numbers contribute to the SNARC effect. Psychonomic Bulletin & Review, 16(2), 328–331. https://doi.org/10.3758/PBR.16.2.328

Shpektor, A., Bakermans, J. J. W., Baram, A. B., Sarnthein, J., Ledergerber, D., Imbach, L., Müller-Seydlitz, E., Barron, H. C., & Behrens, T. E. J. (2024). A hierarchical coordinate system for sequence memory in human entorhinal cortex (p. 2024.10.30.620612). bioRxiv. https://doi.org/10.1101/2024.10.30.620612

Schütt, H. H., Kipnis, A. D., Diedrichsen, J., & Kriegeskorte, N. (2023). Statistical inference on representational geometries. eLife, 12, e82566. https://doi.org/10.7554/eLife.82566

Scoville, W. B., & Milner, B. (1957). Loss of Recent Memory After Bilateral Hippocampal Lesions. Journal of Neurology, Neurosurgery & Psychiatry, 20(1), 11–21. https://doi.org/10.1136/jnnp.20.1.11

Segen, V., Ying, J., Morgan, E., Brandon, M., & Wolbers, T. (2022). Path integration in normal aging and Alzheimer's disease. Trends in Cognitive Sciences, 26(2), 142–158. https://doi.org/10.1016/j.tics.2021.11.001

Shadmehr, R., Brandt, J., & Corkin, S. (1998). Time-Dependent Motor Memory Processes in Amnesic Subjects. Journal of Neurophysiology, 80(3), 1590–1597. https://doi.org/10.1152/jn.1998.80.3.1590

Sharp, P. E., Blair, H. T., & Cho, J. (2001). The anatomical and computational basis of the rat head-direction cell signal. *Trends in Neurosciences*, *24*(5), 289–294. https://doi.org/10.1016/S0166-2236(00)01797-5

Siegel, R. M., & Read, H. L. (1997). Analysis of optic flow in the monkey parietal area 7a. *Cerebral Cortex*, *7*(4), 327–346. https://doi.org/10.1093/cercor/7.4.327

Sigismondi, F. (2024). *Spatial and Conceptual navigation in Early Blind people: Testing the scaffolding hypothesis of cognitive maps*. https://doi.org/10.15168/11572_438691

Sigismondi, F., Xu, Y., Silvestri, M., & Bottini, R. (2024). Altered grid-like coding in early blind people. *Nature Communications*, *15*(1), 3476. https://doi.org/10.1038/s41467-024-47747-x

Skinner, B. F. (1957). *Verbal behavior* (pp. xi, 478). Appleton-Century-Crofts. https://doi.org/10.1037/11256-000

Smith, S. M., & Nichols, T. E. (2009). Threshold-free cluster enhancement: Addressing problems of smoothing, threshold dependence and localisation in cluster inference. NeuroImage, 44(1), 83–98. https://doi.org/10.1016/j.neuroimage.2008.03.061

Solomon, E. A., Lega, B. C., Sperling, M. R., & Kahana, M. J. (2019). Hippocampal theta codes for distances in semantic and temporal spaces. Proceedings of the National Academy of Sciences, 116(48), 24343–24352. https://doi.org/10.1073/pnas.1906729116

Solstad, T., Boccara, C. N., Kropff, E., Moser, M.-B., & Moser, E. I. (2008). Representation of Geometric Borders in the Entorhinal Cortex. *Science*, *322*(5909), 1865–1868. https://doi.org/10.1126/science.1166466

Sorscher, B., Mel, G. C., Ocko, S. A., Giocomo, L. M., & Ganguli, S. (2023). A unified theory for the computational and mechanistic origins of grid cells. *Neuron*, *111*(1), 121-137.e13. https://doi.org/10.1016/j.neuron.2022.10.003

Squire, L. R. (2009). The Legacy of Patient H.M. for Neuroscience. Neuron, 61(1), 6–9. https://doi.org/10.1016/j.neuron.2008.12.023

Stachenfeld, K. L., Botvinick, M. M., & Gershman, S. J. (2017). The hippocampus as a predictive map. *Nature Neuroscience*, *20*(11), Article 11. https://doi.org/10.1038/nn.4650

Staudigl, T., Leszczynski, M., Jacobs, J., Sheth, S. A., Schroeder, C. E., Jensen, O., & Doeller, C. F. (2018). Hexadirectional Modulation of High-Frequency Electrophysiological Activity in the Human Anterior Medial Temporal Lobe Maps Visual Space. Current Biology, 28(20), 3325-3329.e4. https://doi.org/10.1016/j.cub.2018.09.035

Stemmler, M., Mathis, A., & Herz, A. V. M. (2015). Connecting multiple spatial scales to decode the population activity of grid cells. *Science Advances*, *1*(11), e1500816. https://doi.org/10.1126/science.1500816

Summerfield, C. (2022). *Natural General Intelligence: How understanding the brain can help us build AI*. Oxford University Press.

Summerfield, C., Luyckx, F., & Sheahan, H. (2020). Structure learning and the posterior parietal cortex. *Progress in Neurobiology*, *184*, 101717. https://doi.org/10.1016/j.pneurobio.2019.101717

Sun, Y., Nitz, D. A., Xu, X., & Giocomo, L. M. (2024). Subicular neurons encode concave and convex geometries. *Nature*, *627*(8005), 821–829. https://doi.org/10.1038/s41586-024-07139-z

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction*. MIT Press.

Sweigart, S., Nguyen, N., Ranganath, C., Park, S., & Boorman, E. (2023). Cognitive maps at multiple levels of abstraction for flexible inference. *2023 Conference on Cognitive Computational Neuroscience*. 2023 Conference on Cognitive Computational Neuroscience, Oxford, UK. https://doi.org/10.32470/CCN.2023.1283-0

Takahashi, S., Hombe, T., Matsumoto, S., Ide, K., & Yoda, K. (2022). Head direction cells in a migratory bird prefer north. *Science Advances*, *8*(5), eabl6848. https://doi.org/10.1126/sciadv.abl6848

Tarhan, L., & Konkle, T. (2020). Reliability-based voxel selection. NeuroImage, 207, 116350. https://doi.org/10.1016/j.neuroimage.2019.116350

Taube, J. S. (2007). The Head Direction Signal: Origins and Sensory-Motor Integration. *Annual Review of Neuroscience*, *30*(Volume 30, 2007), 181–207. https://doi.org/10.1146/annurev.neuro.29.051605.112854

Taube, J. S., Muller, R. U., & Ranck, J. B. (1990). Head-direction cells recorded from the postsubiculum in freely moving rats. I. Description and quantitative analysis. *Journal of Neuroscience*, *10*(2), 420–435. https://doi.org/10.1523/JNEUROSCI.10-02-00420.1990

Tenderra, R. M., & Theves, S. (2025). Human intelligence relates to neural measures of cognitive map formation. Cell Reports, 44(8). https://doi.org/10.1016/j.celrep.2025.116033

Theves, S., Fernandez, G., & Doeller, C. F. (2019). The Hippocampus Encodes Distances in Multidimensional Feature Space. *Current Biology*, *29*(7), 1226-1231.e3. Scopus. https://doi.org/10.1016/j.cub.2019.02.035

Theves, S., Fernández, G., & Doeller, C. F. (2020). The Hippocampus Maps Concept Space, Not Feature Space. *Journal of Neuroscience*, *40*(38), 7318–7325. https://doi.org/10.1523/JNEUROSCI.0494-20.2020

Tolman, E. C. (1948). Cognitive maps in rats and men. *Psychological Review*, *55*(4), 189–208. https://doi.org/10.1037/h0061626

Tulving, E., Hayman, C. A., & Macdonald, C. A. (1991). Long-lasting perceptual priming and semantic learning in amnesia: A case experiment. Journal of Experimental Psychology: Learning, Memory, and Cognition, 17(4), 595–617. https://doi.org/10.1037/0278-7393.17.4.595

Tustison, N. J., B. B. Avants, P. A. Cook, Y. Zheng, A. Egan, P. A. Yushkevich, and J. C. Gee. 2010. "N4itk: Improved N3 Bias Correction." IEEE Transactions on Medical Imaging 29 (6): 1310–20. https://doi.org/10.1109/TMI.2010.2046908.

Ulanovsky, N., & Moss, C. F. (2007). Hippocampal cellular and network activity in freely moving echolocating bats. *Nature Neuroscience*, *10*(2), 224–233. https://doi.org/10.1038/nn1829

Ulsaker-Janke, I., Waaga, T., Waaga, T., Moser, E. I., & Moser, M.-B. (2023). Grid cells in rats deprived of geometric experience during development. *Proceedings of the National Academy of Sciences*, *120*(41), e2310820120. https://doi.org/10.1073/pnas.2310820120

Urgolites, Z. J., Kim, S., Hopkins, R. O., & Squire, L. R. (2016). Map reading, navigating from maps, and the medial temporal lobe. Proceedings of the National Academy of Sciences, 113(50), 14289–14293. https://doi.org/10.1073/pnas.1617786113

van Ede, F., & Nobre, A. C. (2023). Turning Attention Inside Out: How Working Memory Serves Behavior. *Annual Review of Psychology*, *74*(1), 137–165. https://doi.org/10.1146/annurev-psych-021422-041757

Vericel, M. E., Baraduc, P., Duhamel, J.-R., & Wirth, S. (2024). Organizing space through saccades and fixations between primate posterior parietal cortex and hippocampus. *Nature Communications*, *15*(1), 10448. https://doi.org/10.1038/s41467-024-54736-7

Veselic, S., Muller, T. H., Gutierrez, E., Behrens, T. E. J., Hunt, L. T., Butler, J. L., & Kennerley, S. W. (2025). A cognitive map for value-guided choice in the ventromedial prefrontal cortex. *Cell*, *188*(12), 3259-3273.e22. https://doi.org/10.1016/j.cell.2025.03.038

Viganò, S., & Piazza, M. (2020). Distance and Direction Codes Underlie Navigation of a Novel Semantic Space in the Human Brain. *Journal of Neuroscience*, *40*(13), 2727–2736. https://doi.org/10.1523/JNEUROSCI.1849-19.2020

Viganò, S., & Piazza, M. (2021). The hippocampal-entorhinal system represents nested hierarchical relations between words during concept learning. *Hippocampus*, *31*(6), 557–568. https://doi.org/10.1002/hipo.23320

Viganò, S., Bayramova, R., Doeller, C. F., & Bottini, R. (2023). Mental search of concepts is supported by egocentric vector representations and restructured grid maps. *Nature Communications*, *14*(1), Article 1. https://doi.org/10.1038/s41467-023-43831-w

Viganò, S., Bayramova, R., Doeller, C. F., & Bottini, R. (2024). Spontaneous eye movements reflect the representational geometries of conceptual spaces. *Proceedings of the National Academy of Sciences*, *121*(17), e2403858121. https://doi.org/10.1073/pnas.2403858121

Viganò, S., Giari, G., Mai, R., Doeller, C. F., & Bottini, R. (2025). Foraging in conceptual spaces: Hippocampal oscillatory dynamics underlying searching for concepts in memory (p. 2025.10.07.680852). bioRxiv. https://doi.org/10.1101/2025.10.07.680852

Vollan, A. Z., Gardner, R. J., Moser, M.-B., & Moser, E. I. (2025). Left–right-alternating theta sweeps in entorhinal–hippocampal maps of space. *Nature*, *639*(8056), 995–1005. https://doi.org/10.1038/s41586-024-08527-1

Walther, A., Nili, H., Ejaz, N., Alink, A., Kriegeskorte, N., & Diedrichsen, J. (2016). Reliability of dissimilarity measures for multi-voxel pattern analysis. NeuroImage, 137, 188–200. https://doi.org/10.1016/j.neuroimage.2015.12.012

Walsh, V. (2003). A theory of magnitude: Common cortical metrics of time, space and quantity. Trends in Cognitive Sciences, 7(11), 483–488. https://doi.org/10.1016/j.tics.2003.09.002

Wen, J. H., Sorscher, B., Aery Jones, E. A., Ganguli, S., & Giocomo, L. M. (2024). One-shot entorhinal maps enable flexible navigation in novel environments. *Nature*, *635*(8040), 943–950. https://doi.org/10.1038/s41586-024-08034-3

Whitlock, J. R., Pfuhl, G., Dagslott, N., Moser, M.-B., & Moser, E. I. (2012). Functional Split between Parietal and Entorhinal Cortices in the Rat. *Neuron*, *73*(4), 789–802. https://doi.org/10.1016/j.neuron.2011.12.028

Whitlock, J. R. (2014). Navigating actions through the rodent parietal cortex. Frontiers in Human Neuroscience, 8. https://doi.org/10.3389/fnhum.2014.00293

Whittington, J. C. R., Dorrell, W., Behrens, T. E. J., Ganguli, S., & El-Gaby, M. (2025). A tale of two algorithms: Structured slots explain prefrontal sequence memory and are unified with hippocampal cognitive maps. *Neuron*, *113*(2), 321-333.e6. https://doi.org/10.1016/j.neuron.2024.10.017

Whittington, J. C. R., McCaffary, D., Bakermans, J. J. W., & Behrens, T. E. J. (2022). How to build a cognitive map. *Nature Neuroscience*, *25*(10), Article 10. https://doi.org/10.1038/s41593-022-01153-y

Whittington, J. C. R., Muller, T. H., Mark, S., Chen, G., Barry, C., Burgess, N., & Behrens, T. E. J. (2020). The Tolman-Eichenbaum Machine: Unifying Space and Relational Memory through Generalization in the Hippocampal Formation. *Cell*, *183*(5), 1249-1263.e23. https://doi.org/10.1016/j.cell.2020.10.024

Wilming, N., König, P., König, S., & Buffalo, E. A. (2018). Entorhinal cortex receptive fields are modulated by spatial attention, even without movement. *eLife*, *7*, e31745. https://doi.org/10.7554/eLife.31745

Wood, E. R., Dudchenko, P. A., Robitsek, R. J., & Eichenbaum, H. (2000). Hippocampal Neurons Encode Information about Different Types of Memory Episodes Occurring in the Same Location. *Neuron*, *27*(3), 623–633. https://doi.org/10.1016/S0896-6273(00)00071-4

Yang, C., Mammen, L., Kim, B., Li, M., Robson, D. N., & Li, J. M. (2024). A population code for spatial representation in the zebrafish telencephalon. *Nature*, *634*(8033), 397–406. https://doi.org/10.1038/s41586-024-07867-2

Yang, T., Bavley, R. L., Fomalont, K., Blomstrom, K. J., Mitz, A. R., Turchi, J., Rudebeck, P. H., & Murray, E. A. (2014). Contributions of the hippocampus and entorhinal cortex to rapid visuomotor learning in rhesus monkeys. Hippocampus, 24(9), 1102–1111. https://doi.org/10.1002/hipo.22294

Ying, J., Reboreda, A., Yoshida, M., & Brandon, M. P. (2023). Grid cell disruption in a mouse model of early Alzheimer's disease reflects reduced integration of self-motion cues. *Current Biology*, *0*(0). https://doi.org/10.1016/j.cub.2023.04.065

Yong, E. (2022). *An Immense World: How Animal Senses Reveal the Hidden Realms Around Us: The Sunday Times bestseller*. Random House.

Yu, L. Q., Park, S. A., Sweigart, S. C., Boorman, E. D., & Nassar, M. R. (2021). *Do grid codes afford generalization and flexible decision-making?* (No. arXiv:2106.16219). arXiv. https://doi.org/10.48550/arXiv.2106.16219

Yu, L. Q, Vaidya, A. R., Akhmetzhanova, A. (Aya), Bruinsma, S., & Nassar, M. R. (2025). A parietal grid-like code rotates with cognitive maps but lags rapid behavioral transfer (p. 2025.10.06.680746). bioRxiv. https://doi.org/10.1101/2025.10.06.680746

Zhang, K. (1996). Representation of spatial orientation by the intrinsic dynamics of the head-direction cell ensemble: A theory. *Journal of Neuroscience*, *16*(6), 2112–2126. https://doi.org/10.1523/JNEUROSCI.16-06-02112.1996

Zhang, Y., M. Brady, and S. Smith. 2001. "Segmentation of Brain MR Images Through a Hidden Markov Random Field Model and the Expectation-Maximization Algorithm." IEEE Transactions on Medical Imaging 20 (1): 45–57. https://doi.org/10.1109/42.906424.

Appendix

Additional Details on fMRI Preprocessing

The following boilerplate text is provided by fMRIprep in order to ensure reproducibility, with the express intention that users should copy this information directly into their manuscripts without alteration. Therefore, while basic information about fMRI preprocessing has been provided in the body of the chapter, full details of preprocessing may be found below.

Preprocessing of B0 inhomogeneity mappings

A total of 8 fieldmaps were found available within the input BIDS structure for each subject. A B0-nonuniformity map (or fieldmap) was estimated based on two (or more) echo-planar imaging (EPI) references with topup (Andersson, Skare, and Ashburner (2003); FSL 6.0.5.1:57b01774).

Anatomical data preprocessing

A total of 1 T1-weighted (T1w) images were found within the input BIDS dataset. The T1-weighted (T1w) image was corrected for intensity non-uniformity (INU) with N4BiasFieldCorrection (Tustison et al. 2010), distributed with ANTs 2.3.3 (Avants et al. 2008, RRID:SCR_004757), and used as T1w-reference throughout the workflow. The T1w-reference was then skull-stripped with a Nipype implementation of the antsBrainExtraction.sh workflow (from ANTs), using OASIS30ANTs as target template. Brain tissue segmentation of cerebrospinal fluid (CSF), white-matter (WM) and gray-matter (GM) was performed on the brain-extracted T1w using fast (FSL 6.0.5.1:57b01774, RRID:SCR_002823, Zhang, Brady, and Smith 2001). Volume-based spatial normalization to two standard spaces (MNI152NLin6Asym, MNI152NLin2009cAsym) was performed through nonlinear registration with antsRegistration (ANTs 2.3.3), using brain-extracted versions of both T1w reference and the T1w template. The following templates were selected for spatial normalization: FSL's MNI ICBM 152 non-linear 6th Generation Asymmetric Average Brain Stereotaxic Registration Model [Evans et al. (2012), RRID:SCR_002823; TemplateFlow ID: MNI152NLin6Asym], ICBM 152 Nonlinear Asymmetrical template version 2009c [Fonov et al. (2009), RRID:SCR_008796; TemplateFlow ID: MNI152NLin2009cAsym].

Functional data preprocessing

For each of the 8 BOLD runs found per subject (across all tasks and sessions), the following preprocessing was performed. First, a reference volume and its skull-stripped version were generated using a custom methodology of fMRIPrep. Head-motion parameters with respect to the BOLD reference (transformation matrices, and six corresponding rotation and translation parameters) are estimated before any spatiotemporal filtering using mcflirt (FSL 6.0.5.1:57b01774, Jenkinson et al. 2002). The estimated fieldmap was then aligned with rigid-registration to the target EPI (echo-planar imaging) reference run. The field coefficients were mapped on to the reference EPI using the transform. BOLD runs were slice-time corrected to 0.451s (0.5 of slice acquisition range 0s-0.902s) using 3dTshift from AFNI (Cox and Hyde 1997, RRID:SCR_005927). The BOLD reference was then co-registered to the T1w reference using mri_coreg (FreeSurfer) followed by flirt (FSL 6.0.5.1:57b01774, Jenkinson and Smith 2001) with the boundary-based registration (Greve and Fischl 2009) cost-function. Co-registration was configured with six degrees of freedom. Several confounding time-series were calculated based on the preprocessed BOLD: framewise displacement (FD), DVARS and three region-wise global signals. FD was computed using two formulations following Power (absolute sum of relative motions, Power et al. (2014)) and Jenkinson (relative root mean square displacement between affines, Jenkinson et al. (2002)). FD and DVARS are calculated for each functional run, both using their implementations in Nipype (following the definitions by Power et al. 2014). The three global signals are extracted within the CSF, the WM, and the whole-brain masks. Additionally, a set of physiological regressors were extracted to allow for component-based noise correction (CompCor, Behzadi et al. 2007). Principal components are estimated after high-pass filtering the preprocessed BOLD time-series (using a discrete cosine filter with 128s cut-off) for the two CompCor variants: temporal (tCompCor) and anatomical (aCompCor). tCompCor components are then calculated from the top 2% variable voxels within the brain mask. For aCompCor, three probabilistic masks (CSF, WM and combined CSF+WM) are generated in anatomical space. The implementation differs from that of Behzadi et al. in that instead of eroding the masks by 2 pixels on BOLD space, a mask of pixels that likely contain a volume fraction of GM is subtracted from the aCompCor masks. This mask is obtained by thresholding the corresponding partial volume map at 0.05, and it ensures components are not extracted from voxels containing a minimal fraction of GM. Finally, these masks are resampled into BOLD space and binarized by thresholding at 0.99 (as in the original implementation). Components are also calculated separately within the WM and CSF masks. For each CompCor decomposition, the k components with the largest singular values are retained, such that the retained components' time series are sufficient to explain 50 percent of variance across the nuisance mask (CSF, WM, combined, or temporal). The remaining

components are dropped from consideration. The head-motion estimates calculated in the correction step were also placed within the corresponding confounds file. The confound time series derived from head motion estimates and global signals were expanded with the inclusion of temporal derivatives and quadratic terms for each (Satterthwaite et al. 2013). Frames that exceeded a threshold of 0.5 mm FD or 1.5 standardized DVARS were annotated as motion outliers. Additional nuisance timeseries are calculated by means of principal components analysis of the signal found within a thin band (crown) of voxels around the edge of the brain, as proposed by (Patriat, Reynolds, and Birn 2017). The BOLD time-series were resampled into standard space, generating a preprocessed BOLD run in MNI152NLin6Asym space. First, a reference volume and its skull-stripped version were generated using a custom methodology of fMRIPrep. Automatic removal of motion artefacts using independent component analysis (ICA-AROMA, Pruim et al. 2015) was performed on the preprocessed BOLD on MNI space time-series after removal of non-steady state volumes and spatial smoothing with an isotropic, Gaussian kernel of 6mm FWHM (full-width half-maximum). Corresponding "non-aggressively" denoised runs were produced after such smoothing. Additionally, the "aggressive" noise-regressors were collected and placed in the corresponding confounds file. All resamplings can be performed with a single interpolation step by composing all the pertinent transformations (i.e. head-motion transform matrices, susceptibility distortion correction when available, and co-registrations to anatomical and output spaces). Gridded (volumetric) resamplings were performed using antsApplyTransforms (ANTs), configured with Lanczos interpolation to minimise the smoothing effects of other kernels (Lanczos 1964). Non-gridded (surface) resamplings were performed using mri_vol2surf (FreeSurfer).

Many internal operations of fMRIPrep use Nilearn 0.9.1 (Abraham et al. 2014, RRID:SCR_001362), mostly within the functional processing workflow. For more details of the pipeline, see the section corresponding to workflows in fMRIPrep's documentation.

Results included in this manuscript come from preprocessing performed using fMRIPrep 22.0.1 (Esteban, Markiewicz, et al. (2018); Esteban, Blair, et al. (2018); RRID:SCR_016216), which is based on Nipype 1.8.4 (K. Gorgolewski et al. (2011); K. J. Gorgolewski et al. (2018); RRID:SCR_002502).

Copyright Waiver

The above boilerplate text was automatically generated by fMRIPrep with the express intention that users should copy and paste this text into their manuscripts unchanged. It is released under the CC0 licence.