



Original Articles

Passively learned spatial navigation cues evoke reinforcement learning reward signals

Thomas D. Ferguson^{a,*}, Chad C. Williams^a, Ronald W. Skelton^b, Olave E. Krigolson^a

^a Centre for Biomedical Research, University of Victoria, Canada

^b Department of Psychology, University of Victoria, Canada



ARTICLE INFO

Keywords:

Navigation
Anterior cingulate cortex
Cognitive map
Reward positivity

ABSTRACT

Since the suggestion by Tolman (1948) that both rodents and humans create cognitive maps during navigation, the specifics of how navigators learn about their environment has been mired in debate. One facet of this debate is whether or not the creation of cognitive maps – also known as allocentric navigation – involves reinforcement learning. Here, we demonstrate a role for reinforcement learning during allocentric navigation using event-related brain potentials (ERPs). In the present experiment, participants navigated in a virtual environment that allowed the use of three different navigation strategies (allocentric, egocentric-response, & egocentric-cue), in which their goal was to locate and remember a hidden platform. Following the navigation phase of the experiment, participants were shown “cue images” representative of the three navigation strategies. Specifically, we examined whether or not these passively learned strategy images elicited a reward positivity – an ERP component associated with reinforcement learning and the anterior cingulate cortex. We found that when allocentric navigators were shown previously learned cues predicting the goal location a reward positivity was elicited. The present findings demonstrate that allocentric navigational cues carry long-term value after navigation and lend support to the claim that reinforcement learning plays a role in the acquisition of allocentric navigation and thus the generation of cognitive maps.

1. Introduction

How we learn to navigate space is an area of research wherein there has been, and still is, considerable debate. Initially it was believed that rats could learn to navigate space by building stimulus-response associations (Hull & Spence, 1938). More specifically, the thinking at this point in time was that rats learned via trial and error to use visual cues to navigate to goal locations in mazes. Upon successful navigation, the value of the cues that led to the successful outcome were presumably reinforced via an operant conditioning mechanism (i.e., reinforcement learning: Rescorla & Wagner, 1972). Somewhat problematically for a reinforcement learning account of spatial navigation, subsequent work by Tolman showed that rats were able to select unique paths that they had not previously travelled to find a goal platform (Tolman, 1948). Tolman proposed that rats (and humans) developed cognitive maps that do not get learned as a series of stimulus-response associations but instead that cognitive maps were learned “latently”. In other words, the acquisition of cognitive maps occurs without the animal or human explicitly making specific behavioural responses. Tolman’s discovery led to debate regarding whether or not the learning of cognitive maps is

inherently different from other forms of learning. Specifically, what was in question was whether cognitive maps automatically update (e.g., Doeller & Burgess, 2008; Wang, 2004) or whether they are updated via a reinforcement learning mechanism (e.g., Chamizo, 2003). The general consensus in the literature is that cognitive map learning does follow, on some level, the rules of reinforcement learning (Chamizo, Aznar-Casanova, & Artigas, 2003; Hamilton & Sutherland, 1999; Redhead & Hamilton, 2007; Roberts & Pearce, 1999) given that blocking and overshadowing can impact the learning of spatial cues (Kamin, 1968, 1969). However, the learning of cognitive maps and relative importance of reinforcement learning in this process may also depend on local task goals and how information is presented within an environment (Pearce, 2009; Wang & Brockmole, 2003). In any event, the role of reinforcement learning in cognitive map construction is still not clear.

While the acquisition or use of cognitive maps may be governed in some situations by the rules of reinforcement learning, the observation that cognitive maps can also be learned without agency (i.e., without learning direct stimulus-response associations) suggests that at the very least organisms possess at least two, distinguishable navigation strategies. The two primary navigation strategies that are used are known as

* Corresponding author at: Centre for Biomedical Research, University of Victoria, P.O. Box 1700 STN CSC, Victoria, British Columbia V8W 2Y2, Canada.
E-mail address: tomferg@uvic.ca (T.D. Ferguson).

allocentric (or spatial) and egocentric (or stimulus-response) navigation (Kolb, Sutherland, & Whishaw, 1983). Allocentric navigation, or cognitive map learning (i.e., Tolman, 1948), involves learning the relationships between multiple distal cues in a given environment (O’keefe & Nadel, 1978) and allows navigators to take paths they have never taken before because they possess a cognitive map (as observed in the original Tolman experiments). For example, imagine a university student who is going from class to class who normally follows the same path every day. However, if the student is late for a class, they are able to modify their route without following the same path – perhaps by taking a shortcut through an unexplored area (e.g., cutting across the quad), through the use of previously learned cues that they have incorporated into their cognitive map. Their cognitive map gives them a sense of direction of the location they are heading towards (i.e., the building their class is in), and the imagined student’s ability to take a never before used route is an example of allocentric navigation. This is akin to how some rats in Tolman’s maze were able to navigate to the platform of the Starburst maze using a novel path to get to the rewarded arm, while using cues they had seen before from the surrounding room.

Egocentric navigation, on the other hand, tends to involve the encoding of either one or a series of proximal relationships between stimuli and goal-objects. What we refer to as egocentric navigation has also been called a taxon strategy (O’keefe & Nadel, 1978), non-spatial navigation (Iaria, Petrides, Dagher, Pike, & Bohbot, 2003), or landmark navigation (Foo, Warren, Duchon, & Tarr, 2005). Here, we use the term “egocentric navigation” to specifically refer to stimulus-response navigation, in which participants learn to associate either a body-turn or cue association with a response such as “turn left at the end of the hall” or “head straight to the object”. Within the framework we use here, if participants are learning spatial information with respect to themselves we call this “egocentric” (Klatzky, 1998). Navigators using this strategy are using either a specific object (or scene) to guide their navigation (Trullier, Wiener, Berthoz, & Meyer, 1997). That is, if any aspect of the scene were to change or if the object were to move, then navigators would simply follow that object. Egocentric-cue navigation can either involve learning the association of a goal location with either a visible platform (as in Jacobs, Laurance, & Thomas, 1997) or a proximal cue (as in Jordan, Schadow, Wuestenberg, Heinze, & Jancke, 2004). In support of this view, there is evidence from lesion work in animals that shows that egocentric-cue and egocentric-response both require the caudate nucleus. That is, rats with caudate nucleus damage were unable to find a moving, visible-platform (McDonald & White, 1994), while a similar study found animals with caudate nucleus damage were unable to memorize body-turns to find the platform (Packard & Mcgaugh, 1996).

In contrast to allocentric navigation, egocentric navigation is much quicker (i.e., “see the cue and go”) but does not allow for flexibility. For instance, allocentric navigation allows for the computation of novel paths and shortcuts, as noted previously, but with egocentric navigation the same path must always be taken (see O’keefe & Nadel, 1978 for a more detail). Now imagine a student who is new to campus looking for the library. In order to find the library, the student is told to look for the fountain (a single cue) that sits directly in front of the library building. As such, the student (using an egocentric navigation strategy) would not learn about the location of other buildings and their relative positions but would instead simply learn to use the fountain to find the library. Egocentric navigation strategies have been further subdivided into: (1) cued navigation and (2) response navigation strategies (Etchamendy & Bohbot, 2007; Spriggs, Kirk, & Skelton, 2018). Cued navigation strategies have been defined as the memorization of the relationship between a proximal cue and a goal-object – “the library is directly behind the fountain” (e.g., Hamilton, Kodituwakku, Sutherland, & Savage, 2003). In contrast, response navigation strategies have been defined as the memorization of a single or a series of turns in order to locate a goal-object – “you can get to the library by just going right from where you are” (e.g., Schmitzer-Torbert, 2007).

While the neural systems underlying the acquisition of allocentric navigation strategies are relatively well known (i.e., the role of the hippocampus: O’keefe & Nadel, 1978), far less is known about the neural systems that underlie the acquisition of egocentric navigation strategies. With that said, a study by Baker & Holroyd, 2009 examined the neural basis of the acquisition of egocentric navigation strategies while participants learned to navigate a virtual T-maze¹. In Baker and Holroyd (2009)’s paradigm, participants navigated a series of T-mazes. More specifically, on each experimental trial participants began at the bottom of the “T” in the maze and their choice led them to either the left or right arm following which they received feedback about their choice – an equiprobable outcome of getting a financial reward or punishment. Baker and Holroyd’s results revealed that during performance of the T-maze task both feedback and predictive cues elicited a feedback related negativity (FRN) – a component of the human event-related potential (ERP) evoked by outcome feedback. Interestingly, the FRN is posited to reflect a dopaminergic reinforcement learning signal sent from the basal ganglia to the anterior cingulate cortex (ACC) to optimize behaviour (Holroyd & Coles, 2002) – or in this instance, to learn egocentric navigation cues. Later work also observed a FRN evoked by outcome feedback during navigation of T-mazes (Baker & Holroyd, 2013; Torok et al., 2017).

However, is navigation of Baker and Holroyd’s T-maze actually navigation? For instance, given the lack of movement cues and free movement in the T-maze, both of which play a role in navigation (e.g., Maaswinkel & Whishaw, 1999), it is quite possible that what Baker and Holroyd observed was related to feedback processing rather than any sort of navigational signal. Baker and Holroyd (2009) argued that the FRN was related to spatial navigation due to differences that they observed in the visual N170 ERP component – a result that Torok et al. (2017) were not able to replicate in a later study. With that said, Baker and Holroyd (2013) demonstrated no modulation of the N170 in a task in which feedback was provided against a black background but demonstrated modulation of the N170 when a spatial context was provided. Thus, it appears that the N170 is evoked by some visual aspect of the T-maze. However, it is still unclear whether or not the presence of the FRN in Baker and Holroyd (2009, 2013) and Torok et al. (2017) is related to spatial navigation per se. Indeed, it seems more likely that Baker & Holroyd’s observation of the FRN in the T-maze task is more likely related to the feedback participants received and is not directly tied to the navigational aspect of the task. As such, the specific role of the FRN in navigation remains unclear.

The aforementioned studies using ERP’s to examine the acquisition of navigation strategies by Baker & Holroyd, 2009, 2013 and Torok et al., 2017 suggests three things of importance. Firstly, it suggests that a reinforcement learning system within the medial-frontal cortex (i.e., Holroyd & Coles, 2002) may play some role in navigation strategy acquisition, but the specifics of this role are unclear. Secondly, these studies show the viability of using ERP methods to study navigation. Lastly, these studies highlight the importance of dissociating allocentric and egocentric navigation (as was done in Torok et al., 2017). Of course, if allocentric and egocentric are dissociated, it might also be worth dissociating egocentric-cue and egocentric-response as well.

What remains unclear is the extent of the role of reinforcement learning systems during the development and use of allocentric navigation strategies. As such, here we hoped to further investigate the role of reinforcement learning in navigation in a multi-strategy paradigm

¹ Baker and Holroyd (2009) discussed the development of cognitive maps with regard to their findings. However, given the definitions and literature used here their work would be more correctly termed egocentric navigation (an assertion echoed by Torok et al., 2017) and as such is distinct from allocentric navigation strategies (i.e., cognitive maps). At the very least, it is possible that the lack of dissociation between allocentric and egocentric navigation in the T-maze may have conflated the strategies participants used.

allowing both allocentric and egocentric navigation, by examining how the information that leads navigators to their destination is processed following navigation. We hypothesized that passively learned allocentric cues would evoke reward signals once learned – in a similar manner to what has been observed previously with “active” feedback (i.e., tied to some amount of points) in egocentric navigation (Baker & Holroyd, 2009, 2013). That is, we posited that passively learned allocentric navigation cues (leading to the target location) would evoke a reward positivity² – the positive aspect of the FRN. We hypothesized this because of research showing that cues that predict reward can evoke a reward positivity (Holroyd & Krigolson, 2007; Holroyd, Krigolson, & Lee, 2011; Krigolson, Hassall, & Handy, 2014). The current experiment is the first to examine whether or not the post-navigation presentation of passively learned allocentric navigation cues elicit a reward response. Moreover, based on our reading this is the first experiment to investigate whether or not learned cues continue to elicit a reward positivity following task completion. That is, do cues continue to provide salient information after task completion when no additional feedback is given? Overall, we extend findings from the reinforcement learning literature to navigational research and contribute additional evidence to better understand the role of the reinforcement learning system in navigation.

2. Methods

2.1. Participants

Thirty undergraduate students (14 males, 16 females; mean age 22, 95% CI [20.75, 23.25], age range = 18–31 years) from the University of Victoria took part in the present study and received course credit for their participation. Of the participants, 87% (26/30) were right handed. All participants had normal or corrected to normal vision, no history of brain injury, neurological, or psychiatric complaints, and English as a first language (to minimize misinterpretations of the instructions). All participants provided written, informed consent before beginning the experiment. Ethics approval was obtained from the University of Victoria, Human Ethics Research Board, and all testing was conducted in accordance with ethical standards of the 1964 declaration of Helsinki.

2.2. Materials & apparatus

All testing occurred in a soundproof room, in which participants were seated 12” in front of a 23” monitor (1680 by 1050 pixels). All navigation testing occurred in the HexMaze (Fig. 1A), a modified 6-arm radial arm maze in a virtual environment (see Spriggs et al., 2018 for a more detailed description). Navigation in the Hexmaze allowed for both performance and strategy selection to be measured. The Hexmaze was created using the Unreal Developers Kit (UDK), run on an ordinary laptop (Windows) computer and displayed from a first-person perspective on the monitor, using a 90° field of view. The height of the virtual eyes of the participant was set to be approximately 1.7 m. Participants moved through the environment using an Xbox 360 game controller. Participants turned at a velocity of ± 1.26 rad per second (i.e., it took 5 s to complete a 360° rotation) while the perceived walking speed was ~ 1.5 m/s. The UDK software was modified to permit participants to only move forward, and turn left or right. No backwards movement or strafing was permitted (as might be available in other first-person environments). This was done to mimic the movements available to rats when they were swimming in the original Morris water maze (Morris, 1984). On the controller, both joysticks

² Recently a case has been made that the FRN is actually a reward positivity. A full review of this debate is beyond the scope of this manuscript – see Proudfit (2015) for full details.

were disabled and participants used buttons specifically programmed to travel forward and turn left or right.

2.2.1. Maze environment

The HexMaze was a radial arm maze with 6 arms that was located within a large circular arena. The maze was demarcated by a 1-meter-high arena wall (to prevent participants from leaving the arena during navigation), with a featureless roof (in order to prevent it being used as a clue to orientation), and a tiled floor (which contained a hexagonal portion in the middle tiled in a different direction, to also prevent the floor from being used in orientation). The maze itself was surrounded by a round room with featureless walls. The North and South views each had large windows that presented distal landscape features (a mountain range to the north, an island in water to the south). The eastern and western views had two smaller sets of two windows that presented a view of a slope from the mountains to the water. Of the six arms of the maze, there were two arms that ended facing the walls beside the north and south windows (for a total of four arms; the four directions were: south-east, north-east, south-west, north-west), while to the east and west, each arm ended in the middle of the two smaller windows. At the end of each arm, there were spheres that appeared to “float” above the arena wall. The spheres themselves were either white or blue (the blue sphere was the “cue” sphere that indicated the platform’s location).

2.2.2. Training trials

All participants began the experiment by completing a series of *explore* (n = 1), *visible platform* (n = 4), and *guess* (n = 1) practice trials (see Fig. 1B). Trials were broken up as follows.

2.2.2.1. Explore trials. The purpose of the initial explore practice trial was to ensure participants had sufficient practice with the controller and to familiarize them with the virtual environment. Participants were informed that they could explore the room as much as they wanted.

2.2.2.2. Visible platform trials. Next, participants completed four visible platform practice trials to ensure they could locate a goal location. On these trials participants were instructed to navigate to the visible platform.

2.2.2.3. Guess trials. Finally, participants completed a single guess practice trial. The purpose of the Guess trial was two-fold: (1) to allow participants a chance to get familiar with the Show-Me trial procedures (see below) and (2) to ensure that no aspects of the environment biased the participants to the hidden platform’s upcoming location on Find-It trials (see below). On the Guess trial, participants were instructed to navigate to where they believed the hidden platform might be located.

2.2.3. Learning trials

Following completion of the practice trials all participants completed Find-It (n = 10), and Show-Me (n = 10) learning trials.

2.2.3.1. Find-It trials. Participants completed Find-It trials wherein they located a hidden platform. Each Find-It trial began from the same location (at the end of the south-western arm) and the platform was always located in the eastern arm (that the blue “cue” sphere was above). Participants were instructed to find the platform as quickly and directly as possible, and performance was recorded in the movement time³ and the path distance to find the platform.

³ While movement time is commonly called “latency” or “escape latency” in the navigation literature (e.g., Jacobs et al., 1997; Morris, 1984; Sutherland, Whishaw, & Kolb, 1983) in order to avoid any confusion with ERP component latency, we have chosen instead to refer to navigational latency to find the

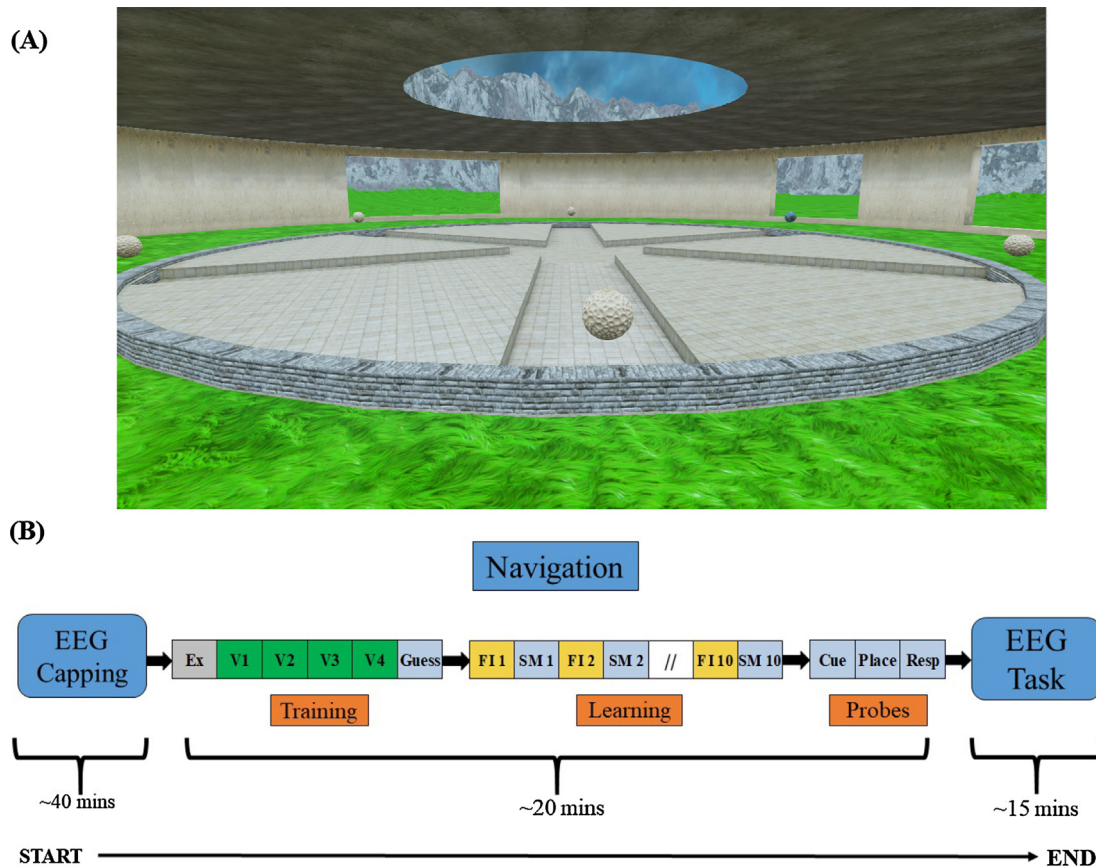


Fig. 1. (A) Screenshot of the HexMaze environment. The floating spheres are seen above the arena wall. The “cue” sphere that showed the platform’s location is indicated in blue (platform not visible). The larger north-facing window, and the two smaller eastern-facing windows are visible. (B) Experimental Protocol from the start to the end of the experiment. Participants were capped, navigated in the maze, and then completed the EEG task. Ex = explore trial, V = visible platform trials, FI = Find-It trials, and SM = Show-Me trials.

2.2.3.2. Show-Me trials. After each Find-It trial, participants completed a Show-Me trial to measure navigation strategy. On Show-Me trials, participants were instructed to navigate to where they thought the platform, which was no longer present, was hidden. Participants began Show-Me trials from the western arm, and the blue sphere that was above the platform on Find-It trials switched locations with a white sphere above another arm. Thus, if participants were navigating by the features of the surrounding landscape (using an *allocentric strategy*), then they would go to the eastern arm (the location of the platform on Find-It trials). If, however, they were using an *egocentric-cue strategy* then they would follow the location of the “cue” sphere, which was shifted to the north-eastern arm. Finally, if they were using an *egocentric-response strategy* (memorizing the long-right body turn), then they would navigate to the south-eastern arm (given the change in their starting position relative to the Find-It trial). Participants completed ten pairs of Find-It and Show-Me trials.

2.2.4. Probe trials

Following completion of the Learning trials, participants completed three strategy probe trials. The purpose of these trials was to determine the extent to which participants’ learned information relating to each strategy individually (allocentric, egocentric cue, egocentric response). In order to isolate whether or not participants learned the strategy, the information relating to one strategy was emphasized and the information of the other two strategies was minimized or removed entirely.

Participants were given identical instructions to the Show-Me trials – to navigate to where they thought the platform was hidden. Participants completed one probe for each of the available strategies for a total of three probes.

2.2.4.1. Place probe trial. On the “Place probe” trial the floating spheres were removed to prevent their use, and the participants began the probe from the center of the maze to prevent body turns being used. Thus, the surrounding landscape was the only available information that could be used by participants to demonstrate their estimate of the platform’s location.

2.2.4.2. Cue probe trial. In contrast, on the “Cue probe”, the windows that allowed for the surrounding landscape to be viewed were blocked and participants began the trial from the center of the maze to prevent the use of body-turns. Moreover, two different colour cues that the participants never saw replaced two of the white spheres. This was done to ensure that participants wouldn’t simply navigate to a unique cue-sphere. Thus, participants could only use the cue spheres to estimate the platform’s location.

2.2.4.3. Response probe trial. On the “Response probe”, the floating cue-spheres were removed, the surrounding landscape was occluded, and participants began a trial from the end of an arm. Thus, participants could only use memorized body-turns to estimate the platform’s location.

2.2.5. EEG task

Following completion of the learning task participants viewed a

(footnote continued)

platform as “movement time”

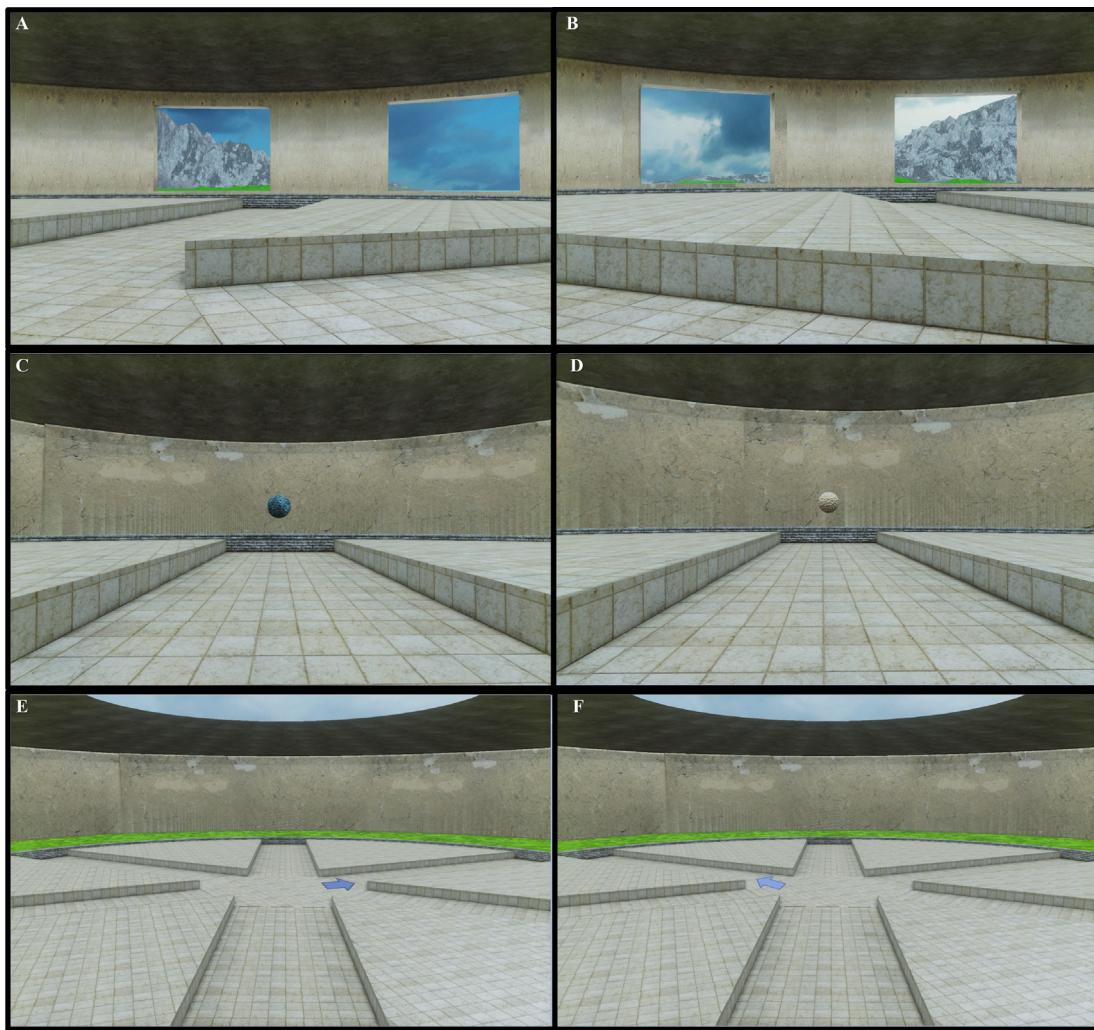


Fig. 2. 6 sample strategy images. (A) Allocentric correct image. (B) Allocentric incorrect image. (C) Cue correct image. (D) Cue incorrect image. (E) Response correct image. (F) Response incorrect image. Note that the arrows in (E) and (F) were not present during navigation in the HexMaze.

series of “strategy” images while EEG data were recorded. These “strategy” images (allocentric, egocentric-cue, egocentric-response) were presented in MATLAB (Version 8.6, Mathworks, Natick, U.S.A.) using the Psychophysics Toolbox extension (Brainard, 1997; Pelli, 1997). Specifically, 36 different view point images were presented, with each image being a screenshot taken from a different point within the Hexmaze. The 36 images were sub-divided into three categories reflecting the three strategies used by participants (allocentric, egocentric-cue, egocentric-response: 12 images per strategy; see Fig. 2 for example images). In a key manipulation, within each set of the 12 images for each strategy 6 images were “correct” images and 6 images were “incorrect” images. The 6 correct images corresponded to the correct platform location. The 6 “incorrect” images were images that had the same information as the correct images, but instead led to a location where there had been no platform.

The images were presented in 15 blocks of 36 trials. Immediately before viewing each image, a fixation cross in the center of the screen was presented for 400–600 ms followed by a randomly presented strategy image for 400–600 ms. Participants saw each of the 36 images 15 times each, and the individual images were presented in a completely randomized order. Additionally, on random trials a white asterisk was shown superimposed on the presented image at a rate of around 1 out of every 36 trials to ensure participants were paying attention to the presented images (c.f., Handy, Tipper, Borg, Grafton, & Gazzaniga, 2006). Participants were instructed to click the left mouse

button in order to move on from these “white-asterisk” trials. Following each block, participants were given a self-paced rest period. At the conclusion of the viewing task, participants completed a brief post-test questionnaire. The purpose of this was to identify confounds such as whether or not they played video-games regularly (and what type – e.g., 3D, 2D), whether they had consumed caffeine, how much sleep they had gotten the night before, and if they found their attention wandering at any point during the task.

2.3. Data acquisition

In the Hex-maze, our experimental software recorded movement time and path distance on the Find-It trials as measures of performance. Strategy choices on Show-Me trials and the strategy probes were recorded as measures of X and Y coordinates in the HexMaze. As well, in the EEG image presentation task, reaction time was recorded on the white asterisks trials (that is, how long it took them to click). Electroencephalographic (EEG) data were recorded from 64 electrodes that were mounted in a fitted cap with a standard 10–20 layout (ActiCAP, Brain Products GmbH, Munich, Germany) were recorded using Brain Vision Recorder software (Version 1.10, Brain Products GmbH, Munich, Germany). All electrodes were referenced to a common ground and during recording electrode impedances were kept below 20 k Ω . EEG data were sampled at 500 Hz, amplified (ActiCHamp, Revision 2, Brain Products GmbH, Munich, Germany), and filtered through an

antialiasing low-pass filter of 8 kHz.

2.4. Data analysis

2.4.1. Behavioural

To determine strategy selection (of the three strategies: allocentric, egocentric-cue, & egocentric-response), we classified participants using their strategy choice on Show-Me trials, which was determined by which arm participants chose on those trials. In other words, if the participant went to the location of the platform relative to the surrounding landscape then we classified this as the selection of an allocentric strategy on that trial. However, if the participant followed the blue cue-sphere we classified this as the selection of an egocentric-cue strategy. Finally, if the participant made the same response (i.e., body-turn) as on the Find-It trial then we classified this as the selection of an egocentric-response strategy. Across trials, if participants selected one of the strategies more often than the other strategies (and if they selected that one strategy on more than 4 trials in total), then they were classified as having used that strategy. In terms of behavioural measures in the HexMaze, we computed means for both movement time (in seconds) and path distance (in pool diameters, a measure of path length calculated as a proportion of the distance travelled divided by the total diameter of the maze) across all 10 Find-It trials. For our error measurements, we computed 95% confidence intervals (Loftus & Masson, 1994). To assess whether participants improved in performance across the task we utilized a nested model comparison approach. Specifically, for our two measures of performance – movement time and path distance – we compared a uniform baseline model (intercept-only) to an alternative model (logarithmic) and compared the models with a chi-square goodness of fit test. Significance levels for all tests were set at $\alpha = 0.05$. On the EEG task, we computed the average reaction times (in ms) to respond to the white asterisk. No participants had average reaction times to respond greater than 2 standard deviations from the mean, and as such, no participants were excluded from analysis. All statistical tests were conducted in R (version 3.4.3; R Core Team, 2017).

2.4.2. EEG

EEG data were processed using standard methods in the Krigolson Laboratory (<http://www.krigolsonlab.com/data-analysis.html>). EEG data were recorded at 500 Hz from 64 electrodes spread out across the scalp. All excessively noisy and faulty channels were removed from analysis at the outset (although later reinterpolated). The EEG data were then down-sampled to 250 Hz, and all EEG data were re-referenced to an average of the two mastoid electrodes (TP9, TP10). Following this, data were filtered using a dual-pass Butterworth filter with a passband of 0.1–30 Hz, and a notch filter was applied at 60 Hz. Segments were then created from –1000 ms to 2000 ms centred on each event of interest (the strategy image that was shown) from the continuous EEG. Following this, independent component analysis (ICA) was conducted to identify ocular artifacts (Luck, 2014) and the corresponding ICA components were removed. Then data were reconstructed using the remaining ICA components and any removed channels were interpolated using the method of spherical splines. To examine the ERP response to the visual cue stimuli, the data were segmented again using a shorter epoch (–200 ms to 600 ms) again around each event of interest. All segments were baseline corrected using a 200 ms window preceding stimuli onset. Finally, all segments underwent an artifact rejection algorithm that removed segments that had gradients greater than 10 $\mu\text{V}/\text{ms}$ and/or a 100 μV absolute within-segment difference. The artifact rejection algorithm led to an average rejection of 14.7%, (95% CI [11.2%, 18.2%]) of the total EEG data for each participant. Following artifact rejection participant average ERP waveforms were computed for each of the six strategy image types: (1) allocentric correct, (2) allocentric incorrect, (3) egocentric-cue correct, (4) egocentric-cue incorrect, (5) egocentric-response correct, & (6) egocentric cue incorrect). Additionally, difference waveforms were created by

subtracting the incorrect location from the correct location cue images for each strategy type. Finally, grand average conditional and difference waveforms were computed from the individual averaged data.

2.4.2.1. Reward positivity. The sole ERP component of interest in this study was the reward positivity – a frontal-central ERP component that reaches maximum 250–350 ms following the presentation of feedback (e.g., Miltner, Braun, & Coles, 1997; Proudfit, 2015), although feedback in this case was the strategy image. Based on prior research (Holroyd & Coles, 2002; Holroyd & Krigolson, 2007) and topographic inspection of the data, we quantified the reward positivity at channel Cz (where it was maximal) on a participant-by-participant basis by calculating the mean voltage \pm 40 ms of the maximal difference on the grand average waveform (336 ms). We computed the reward positivity using the difference wave approach for each strategy category (taking the difference between correct strategy images and their corresponding incorrect strategy image). We conducted single-sample *t*-tests ($\alpha = 0.05$) of these difference wave peaks against zero (Holroyd & Krigolson, 2007) to verify the existence of the reward positivity. For our error measurements, we again computed 95% confidence intervals, and Cohen's *d* was computed for all *t*-tests.

2.4.2.2. Cluster analysis. We also implemented a non-parametric cluster permutation using FieldTrip (Oostenveld, Fries, Maris, & Schoffelen, 2011) to control for the multiple comparison correction while allowing us to avoid making strong assumptions regarding effect location or timing. As per the ERP analysis, we used this analysis to compare the average ERP responses. In the cluster analysis, the average ERP responses were compared across all electrodes between 200 and 400 ms post presentation of the stimuli. Only one adjacent significant cluster (either temporally or spatially) was needed for a cluster to be significant. This analysis was repeated for all three image types (allocentric, egocentric-cue, and egocentric-response).

2.4.2.3. Spatial temporal PCA. In order to verify the existence, topography, and timing of the reward-positivity we conducted a spatial-temporal PCA with Varimax rotation using the ERP-PCA toolkit (Dien, 2010) on our average ERP waveforms for each condition, similar to Spencer, Dien, & Donchin, 2001. In brief, first we submitted the spatial data (channels) to the PCA analysis. Then, we examined the component with maximal factor loadings (i.e., the spatial component that explained the most of the variance) and submitted the virtual ERPs for this component to a temporal PCA to examine component timing (using the temporal component that explained the most of the variance). We choose a Varimax rotation for the PCA analysis because it maximizes the variance of the sub-space to the smallest amount of components (c.f. Donchin & Heffley, 1978). Finally, we computed spatial temporal PCA scores for each of the conditions across all 22 allocentric navigators and compared them using a paired *t*-test (with Cohen's *d* as the measure of effect size and 95% confidence intervals as the measure of error).

3. Results

3.1. Behavioural analysis

In the navigation task, the vast majority of navigators selected an allocentric strategy. That is, out of the 30 total navigators, most navigators choose to navigate allocentrically (22/30), which meant that we only had enough participants to analyze allocentric navigators. Thus, both the egocentric-cue (5/30) and the egocentric-response (3/30) navigators were excluded from any further statistical analysis in the results – however, see the [supplementary materials](#) section for an exploratory analysis of the egocentric navigators.

Analysis of participants' performance in the navigation task revealed that participants did improve in performance across trials

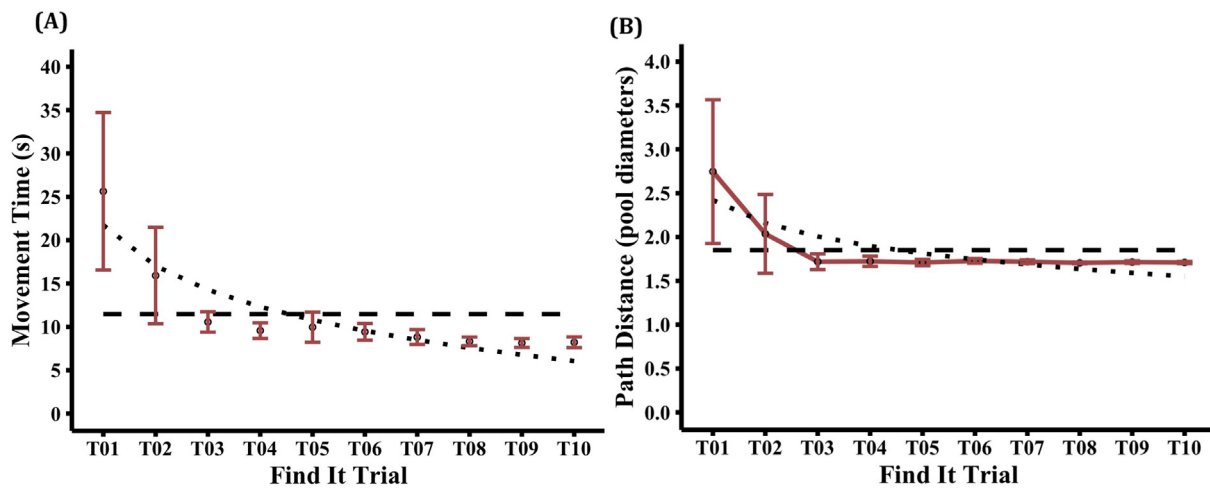


Fig. 3. Average performance (of allocentric navigators) across all 10 Find-It trials. (A) Movement time to find the platform. (B) Path distance travelled to find the platform. The dashed line indicates the uniform-curve fit to the data (the baseline model), while the dotted line indicates a log-curve fit to the data (the full model). Error bars represent 95% confidence intervals.

(Fig. 3). Specifically, for movement time on Find-It trials we found that a logarithmic model fit the data better than the uniform model, $\chi^2(1) = 4934.90$, $p < .001$, demonstrating that movement time decreased and then plateaued. Our analysis of path distance paralleled this result, as the logarithmic model fit the data significantly better than the uniform model, $\chi^2(1) = 15.23$, $p < .001$, highlighting a decrease and then plateau in path distance with practice. These findings from both movement time and path distance suggest that while performance improved over the ten trials, this improvement leveled off during the final Find-It trials – that is, participants reached asymptote. Visual examination suggests that this was the case after trial 3 for both movement time and path distance.

3.2. EEG analysis

3.2.1. Reward positivity

A visual inspection of the ERP data demonstrated a component with the timing and topography consistent with the reward positivity for the allocentric images, but not for the cue and response images (see Fig. 4).

Our statistical analysis confirmed this – after stimulus presentation of the allocentric images, a reward positivity was present with an average deflection of $1.20 \mu\text{V}$ at Cz, which was different from zero, $t(21) = 3.63$, $p < .005$, $d = 0.79$, 95% CI [$+0.51 + 1.88$]. There was, however, no reward positivity at Cz for the egocentric-cue images (average deflection of $0.60 \mu\text{V}$, $t(21) = 1.17$, $p = .25$, $d = 0.26$, 95% CI [$-0.46 + 1.66$]) nor the egocentric-response images, (average deflection of $0.20 \mu\text{V}$, $t(21) = 0.46$, $p = .65$, $d = 0.10$, 95% CI [$-0.65 + 1.02$]).

A non-parametric cluster analysis confirmed the ERP analysis – differences were observed between correct and incorrect waveforms for allocentric images. Specifically, for allocentric images we observed a cluster ($p < .05$) that included electrodes FC2, Cz, C2, and CP2 between 300 and 400 ms. For both response and cue images no clusters were observed (all p 's > 0.10).

3.2.2. PCA analysis

Results of the spatial-temporal PCA analysis confirmed that our identified component was a reward positivity. The results from the spatial PCA (Fig. 5 – panel A) demonstrated that, a frontal-central spatial factor accounted for the most variance (48.6%). The spatial PCA analysis itself only yielded two spatial factors that accounted for greater than 5% of variance, and the second factor was a posterior spatial factor accounting for 32.9% of the variance (thus, the two components together accounted for 81.5% of the total variance). The results from the

temporal PCA suggested four factors accounting for greater than 5% of variance (see Fig. 5 – panel B). Of these temporal factors, the first factor matched the timing of our component of interest and explained the most variance (28.3%). Thus, this factor was selected for further analysis and for plotting for the spatial temporal scores (see Fig. 5 – panel C). Results from the paired t -test of the spatial temporal scores showed that the scores from the allocentric correct images ($M = -1.39$) were more positive than the scores from the allocentric incorrect images ($M = -1.71$, $M_{\text{diff}} = 0.32$), $t(21) = 2.36$, $p < .05$, $d = 0.32$, 95% CI [$+0.04 + 0.59$]).

4. Discussion

In the present experiment we examined whether or not passively learned spatial navigation cues evoked reward responses in order to further elucidate the role reinforcement learning plays in allocentric spatial navigation. Our behavioural data demonstrated that the majority of participants selected an allocentric as opposed to an egocentric navigation strategy. Furthermore, our behavioural data demonstrated that the allocentric navigators learned to navigate the HexMaze given the observed reductions in movement time and path distance. In terms of our ERP results collected following task completion, we observed a frontal-central positive deflection for correct relative to incorrect spatial cues – a reward positivity. This result was confirmed with a non-parametric cluster analysis and a spatial-temporal PCA analysis. Interestingly, this same frontal-central positive deflection was not present when allocentric navigators viewed egocentric-cue or egocentric-response images suggesting there was something particular about the allocentric images that elicited a reward response.

The results presented here suggest that to some extent allocentric navigators relied upon a reinforcement learning mechanism within the medial-frontal cortex (c.f., Holroyd & Coles, 2002) to strengthen the value of spatial cues within the HexMaze. Although the spatial cues do not reflect rewards per se, they are predictive of the location of the task goal and as such should acquire value via a reinforcement learning process (i.e., Sutton & Barto, 1998). Our work is in line with previous research showing that predictive cues acquire value with learning and elicit a reward positivity (Baker & Holroyd, 2009; Holroyd & Krigolson, 2007; Holroyd et al., 2011; Krigolson et al., 2014). Our results are also in line with seminal studies in monkeys that have shown that dopamine neurons increase their firing in response to predictive cues following learning (Schultz, Dayan, & Montague, 1997). Further to this point, previous physiological work understanding the dopamine pathways in rats further suggests a role of the dopamine system in allocentric

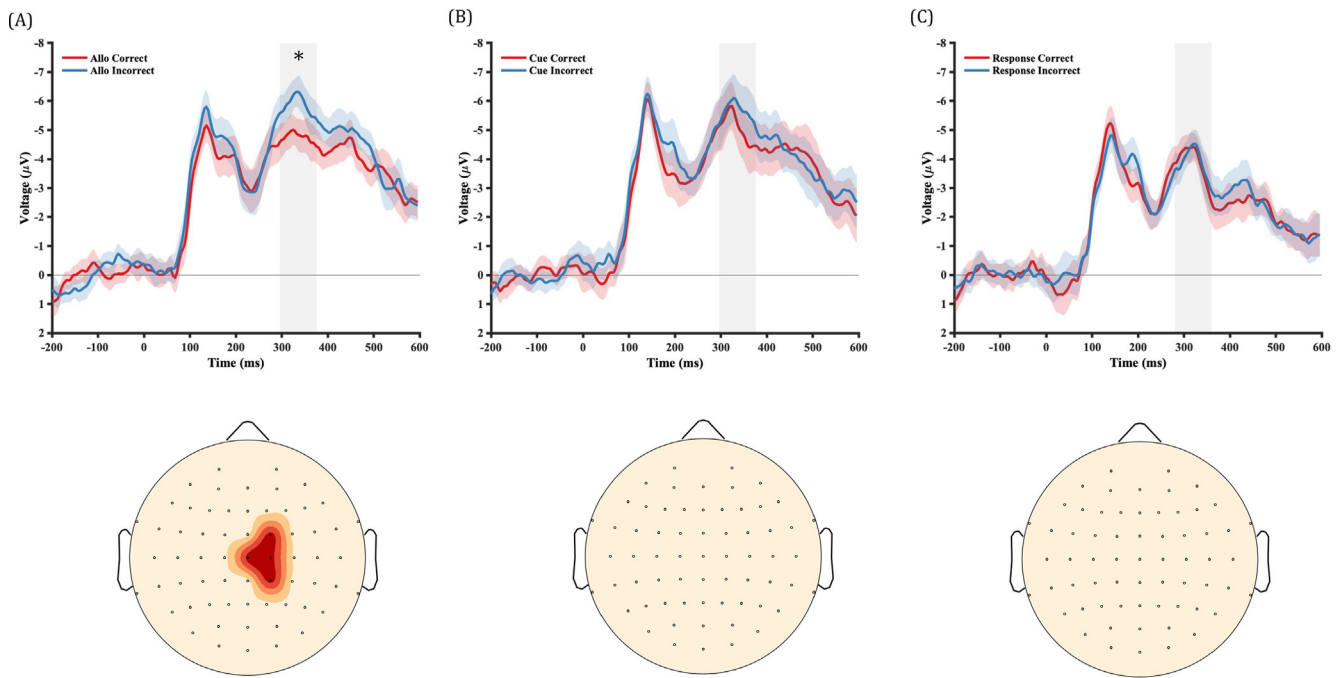


Fig. 4. Time-locked neural waveforms. (A) Neural waveforms for both allocentric correct (indicated in red) and allocentric incorrect (indicated in blue) images, centered around 336 ms (from 296 ms to 376 ms, window indicated in gray). (B) Neural waveforms for egocentric-cue correct (indicated in red) and egocentric-cue incorrect (indicated in blue) images. (C) Neural waveforms for egocentric-response correct (indicated in red) and egocentric-response incorrect (indicated in blue) images. Topographic maps show the significant electrodes for each image category as per the cluster analysis, indicated in red. All waveform data are taken from electrode Cz. Error-bars indicate 95% confidence intervals. Asterisk (*) indicates a significant difference.

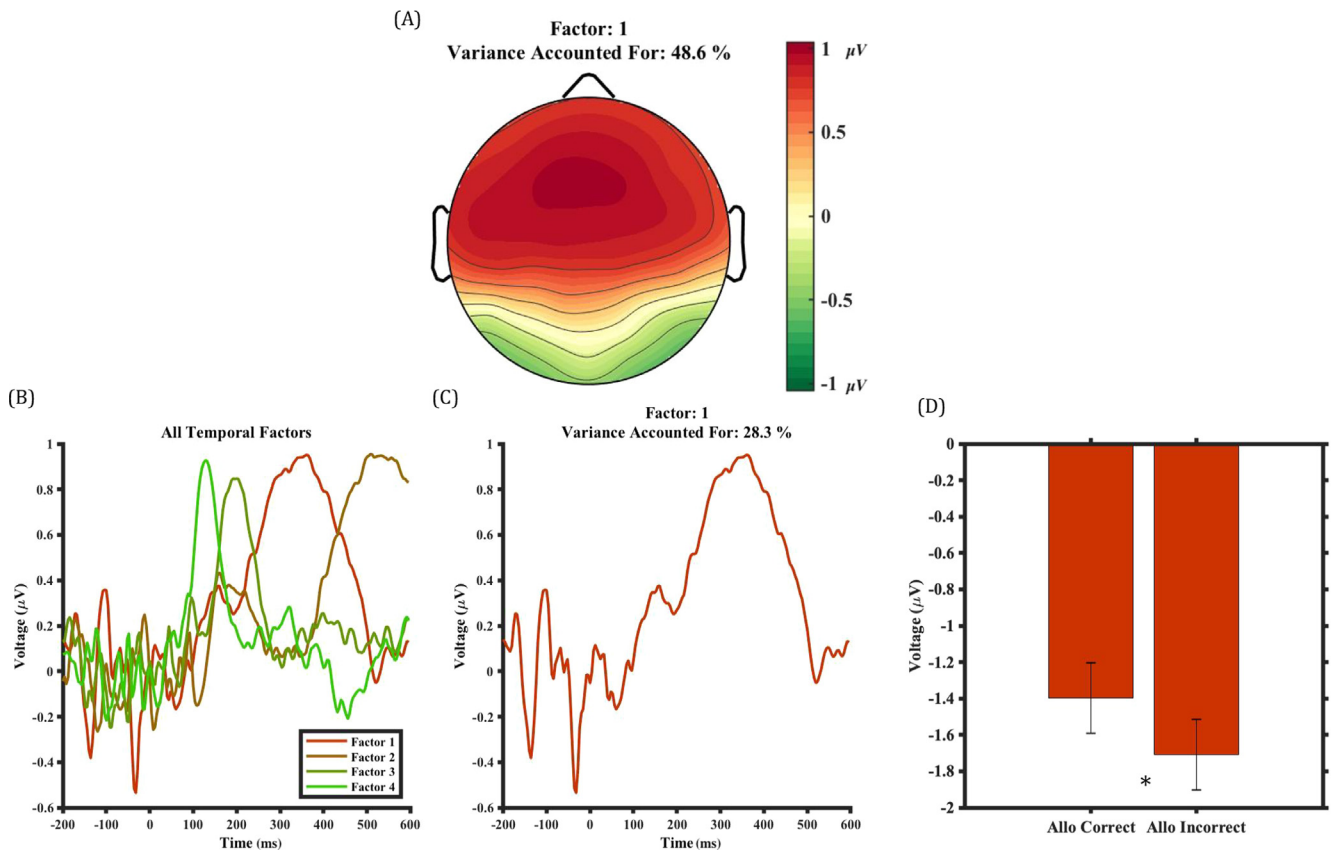


Fig. 5. Results from the Spatial-Temporal PCA. (A) The topographical map of the factor loading that contributed the most variance from the Spatial PCA. (B) The grand average waveforms from the resulting Temporal PCA conducted on the spatial factor loading from A (left) and an isolation of waveform that contributed to the most variance (right). (C) The resultant spatial-temporal PCA scores of the component isolated in B. Error bars are 95% confidence intervals. Asterisk (*) indicates a significant difference.

navigation (Braun, Graham, Schaefer, Vorhees, & Williams, 2012). That is, the authors found that the dopaminergic-striatal pathway was involved in both egocentric and allocentric navigation, as rats injected with a neurotoxic compound that destroys or depletes dopamine neurons were impaired in both types of navigation (separate of any motor or motivational deficits), showing allocentric navigation is dependent on the dopaminergic system to some extent. Our results also show that cues have long-term values that, depending on the context, may continue to be salient past completion of the task. This finding is entirely novel, but given the uniqueness of cognitive map learning (e.g., learning in the hippocampus is uniquely fast compared to learning in the rest of the neo-cortex – McClelland, McNaughton, & O'Reilly, 1995), it is important to determine whether or not this is unique to cognitive map navigation or whether this is the case for all reinforcement learning paradigms.

Interpreting the present work in light of previous work investigating the role of reinforcement learning in spatial navigation is important. Baker and Holroyd (2009, 2013) do make a compelling claim that the principles of reinforcement learning govern some aspect of cognitive map navigation – they observed that rewards and associated predictive cues elicited a reward positivity during navigation of a “T-Maze”. However, given the structure of the “T-Maze” task it seems clear that spatial navigation during performance of this task is egocentric as opposed to allocentric navigation (if it is navigationally related at all – see our discussion of this point in the introduction). It is also worth noting here that Baker and Holroyd provided formal feedback (i.e., a symbol appeared indicating task outcome: as did Torok et al., 2017). Specifically, Baker and Holroyd and Torok et al. provided feedback that was tied to some amount of points – or not in the case of the no-reward feedback – and a monetary bonus was tied to their point total after each learning trial. Here, we simply showed participants cues that either indicated the correct or incorrect location of the navigation goal and demonstrated that these passively learned cues elicited a reward positivity.

Overall, the current findings provide further evidence that reinforcement learning plays a role in allocentric navigation. Indeed, while previous research suggests that the principals of reinforcement learning play some role in allocentric navigation (e.g., Pearce, 2009; Hamilton & Sutherland, 1999), the present findings directly confirm involvement of the medial-frontal reinforcement learning system (i.e., Holroyd & Coles, 2002) in valuating allocentric navigational information. Given the need for planning during navigation – and given that planning is central to the frontal cortex (e.g., Karnath, Wallesch, & Zimmermann, 1991) – it makes sense that there is involvement of the frontal lobes during navigation. Moreover, better allocentric navigation performance is correlated with higher performance levels on measures of executive function (such as set switching and response inhibition; Korthauer, Nowak, Frahm, & Driscoll, 2017), linking navigational abilities and executive function abilities. In fact, neuroimaging research using EEG has shown that the frontal lobes are active during goal seeking while navigating (Caplan et al., 2003), while findings from fMRI research has shown that the pre-frontal cortex is active in taxi drivers both during navigational planning and when switching routes (Spiers & Maguire, 2006). Thus, the present work extends these findings, in that it confirms the involvement of medial-frontal cortex not just during navigational planning and goal-seeking, but also when passively viewing navigational cues following navigation.

Assuming the reward positivity reflects a reinforcement learning signal generated within the ACC (e.g., Holroyd & Coles, 2002), our findings also suggest a role for the ACC in allocentric navigation. This matches up with physiological evidence from a lesion study in rodents⁴

which showed that deficits in navigational abilities were observed in rats given ACC lesions (Sutherland, Whishaw, & Kolb, 1988). However, what role the ACC plays in navigation remains unclear. Perhaps the ACC facilitates navigators deciding between strategies in a hierarchical manner. For instance, it has been suggested that the reward positivity reflects learning of the value of a task itself but not the learning of values for individual trials (Umehoto, HajiHosseini, Yates, & Holroyd, 2017) suggesting a higher-level role of the ACC in task monitoring via hierarchical reinforcement (e.g., Holroyd & McClure, 2015; Holroyd & Yeung, 2012). Thus, perhaps the ACC monitors the current environment in order to optimize behaviour and switch navigation strategies as necessary. This could explain why some research has shown allocentric to egocentric (Iaria et al., 2003) or egocentric to allocentric (Harris & Wolbers, 2014) or bi-directional (Iglói, Zaoui, Berthoz, & Rondi-Reig, 2009) switching of navigation strategies. That is, depending on the environmental demands, the ACC chooses the most appropriate navigation strategy. Alternatively, given a possible role of the ACC in resource allocation (e.g., Vassena, Holroyd, & Alexander, 2017), it could also be that the ACC is active to ensure that we are learning optimal behavioural strategies in order to be as efficient as possible. For example, if the ACC detects it would be better to use a simpler egocentric strategy rather than continuing to use a more complex allocentric strategy, then it may send signals to the navigator to switch strategies.

One additional point to consider given the data presented here relates to the asymptote in performance observed following trial three. Indeed, our behavioural data suggest that no further learning occurred after this point. Importantly, this lack of learning could potentially influence the amplitude of the reward positivity; for example, Krigolson et al., 2014 reported a reduction in the amplitude of the reward positivity following learning. Thus, if we had stopped the learning phase in the present experiment after trial three we might predict a slightly larger reward positivity compared to the reward positivity we found following completion of all of the learning trials. There are two reasons for this. The first reason is that as participants become more certain as they learn a task the amplitude of their reward positivity decreases (e.g., Eppinger, Kray, Mock, & Mecklinger, 2008). The second reason relates to the belief that the reward positivity reflects a reinforcement learning prediction error that is sensitive to whether or not outcomes are better or worse than expected (Holroyd et al., 2011). Within this framework, in a deterministic-task such as ours, the amplitude of the reward positivity would be reduced with learning as prediction errors diminish in amplitude.

4.1. Conclusions

Taken together, the present findings demonstrate that passively learned cues acquire value via a reinforcement learning mechanism during allocentric spatial learning. Specifically, we demonstrated that passively learned spatial cues elicited a reward positivity if they indicated the correct location of a navigation goal – to the best of our knowledge a novel result highlighting the role of reinforcement learning in the acquisition of allocentric navigation. As well, these passively learned navigational cues have long-term value. However, further elucidation is needed to determine just how much of a role the reinforcement learning system plays in allocentric navigation, and what specifically is the role of the ACC. Does the ACC contribute in any way to strategy selection, or through the hierarchical control of navigation strategies directly?

Funding

This work was supported by the National Science and Engineering Research Council of Canada [Grant Number RGPIN 2016-0943].

⁴ There may be differences in the structural layout of the ACC in rats and the ACC in humans (see Vogt, 2016 for more details), and thus some caution should be taken when interpreting these results.

Declarations of interest

None.

Appendix A. Supplementary material

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.cognition.2019.03.015>. The link to the data itself is: <https://osf.io/z6hnp/>. The DOI link to the database is: <https://doi.org/10.17605/OSF.IO/Z6HNP>.

References

- Baker, T. E., & Holroyd, C. B. (2009). Which way do I go? Neural activation in response to feedback and spatial processing in a virtual T-maze. *Cerebral Cortex*, 19(8), 1708–1722. <https://doi.org/10.1093/cercor/bhn223>.
- Baker, T. E., & Holroyd, C. B. (2013). The topographical N170: Electrophysiological evidence of a neural mechanism for human spatial navigation. *Biological Psychology*, 94(1), 90–105. <https://doi.org/10.1016/j.biopsycho.2013.05.004>.
- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, 10(4), 433–436.
- Braun, A. A., Graham, D. L., Schaefer, T. L., Vorhees, C. V., & Williams, M. T. (2012). Dorsal striatal dopamine depletion impairs both allocentric and egocentric navigation in rats. *Neurobiology of Learning and Memory*, 97(4), 402–408. <https://doi.org/10.1016/j.nlm.2012.03.004>.
- Caplan, J. B., Madsen, J. R., Schulze-Bonhage, A., Aschenbrenner-Scheibe, R., Newman, E. L., & Kahana, M. J. (2003). Human theta oscillations related to sensorimotor integration and spatial learning. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 23(11), 4726–4736.
- Chamizo, V. D. (2003). Assessing the generality of the mechanism of learning. *The Quarterly Journal of Experimental Psychology: Section B*, 56(1), 102–113. <https://doi.org/10.1080/02724990244000205>.
- Chamizo, V. D., Aznar-Casanova, J. A., & Artigas, A. A. (2003). Human overshadowing in a virtual pool: Simple guidance is a good competitor against locale learning. *Learning and Motivation*, 34(3), 262–281. [https://doi.org/10.1016/S0023-9690\(03\)00020-1](https://doi.org/10.1016/S0023-9690(03)00020-1).
- Dien, J. (2010). The ERP PCA Toolkit: An open source program for advanced statistical analysis of event-related potential data. *Journal of Neuroscience Methods*, 187(1), 138–145. <https://doi.org/10.1016/j.jneumeth.2009.12.009>.
- Doeller, C. F., & Burgess, N. (2008). Distinct error-correcting and incidental learning of location relative to landmarks and boundaries. *Proceedings of the National Academy of Sciences*, 105(15), 5909–5914. <https://doi.org/10.1073/pnas.0711433105>.
- Donchin, E., & Heffley, E. (1978). Multivariate analysis of event-related potential data: a tutorial review. In D. Otto (Ed.), *Multidisciplinary perspectives in event-related potential research* (pp. 555–572). Washington, D.C.: U.S. Government Printing Office.
- Eppinger, B., Kray, J., Mock, B., & Mecklinger, A. (2008). Better or worse than expected? Aging, learning, and the ERN. *Neuropsychologia*, 46, 521–539. <https://doi.org/10.1016/j.neuropsychologia.2007.09.001>.
- Etchamendy, N., & Bohbot, V. D. (2007). Spontaneous navigational strategies and performance in the virtual town. *Hippocampus*, 17, 595–599.
- Foo, P., Warren, W. H., Duchon, A., & Tarr, M. J. (2005). Do humans integrate routes into a cognitive map? Map-versus landmark-based navigation of novel shortcuts. *Journal of Experimental Psychology*, 31(2), 195–215. <https://doi.org/10.1037/0278-7393.31.2.195>.
- Hamilton, D. A., Kodituwakku, P., Sutherland, R. J., & Savage, D. D. (2003). Children with fetal alcohol syndrome are impaired at place learning but not cued-navigation in a virtual Morris water task. *Behavioural Brain Research*, 143, 85–94. [https://doi.org/10.1016/S0166-4328\(03\)00028-7](https://doi.org/10.1016/S0166-4328(03)00028-7).
- Hamilton, D. A., & Sutherland, R. J. (1999). Blocking in human place learning: Evidence from virtual navigation. *Psychobiology*, 27(4), 453–461.
- Handy, T. C., Tipper, C. M., Borg, J. S., Grafton, S. T., & Gazzaniga, M. S. (2006). Motor experience with graspable objects reduces their implicit analysis in visual- and motor-related cortex. *Brain Research*, 1097(1), 156–166. <https://doi.org/10.1016/j.brainres.2006.04.059>.
- Harris, M. A., & Wolbers, T. (2014). How age-related strategy switching deficits affect wayfinding in complex environments. *Neurobiology of Aging*, 35(5), 1095–1102. <https://doi.org/10.1016/j.neurobiolaging.2013.10.086>.
- Holroyd, C. B., & Coles, M. G. H. (2002). The neural basis of human error processing: Reinforcement learning, dopamine, and the error-related negativity. *Psychological Review*, 109(4), 679–709. <https://doi.org/10.1037/0033-295X.109.4.679>.
- Holroyd, C. B., & Krigolson, O. E. (2007). Reward prediction error signals associated with a modified time estimation task. *Psychophysiology*, 44, 913–917. <https://doi.org/10.1111/j.1469-8986.2007.00561.x>.
- Holroyd, C. B., Krigolson, O. E., & Lee, S. (2011). Reward positivity elicited by predictive cues. *NeuroReport*, 22(5), 249–252. <https://doi.org/10.1097/WNR.0b013e328345441d>.
- Holroyd, C. B., & McClure, S. M. (2015). Hierarchical control over effortful behavior by rodent medial frontal cortex: A computational model. *Psychology Review*, 122(1), 54–83. <https://doi.org/10.1037/a0038339>.
- Holroyd, C. B., & Yeung, N. (2012). Motivation of extended behaviors by anterior cingulate cortex. *Trends in Cognitive Sciences*, 16(2), 122–128. <https://doi.org/10.1016/J.TICS.2011.12.008>.
- Hull, C., & Spence, K. (1998). ‘Correction’ vs. ‘non-correction’ method of trial-and-error learning in rats. *Journal of Comparative Psychology*, 25, 127.
- Iaria, G., Petrides, M., Dagher, A., Pike, B., & Bohbot, V. D. (2003). Cognitive strategies dependent on the hippocampus and caudate nucleus in human navigation: Variability and change with practice. *Journal of Neuroscience*, 23(13), 5945–5952.
- Igló, K., Zaoui, M., Berthoz, A., & Rondi-Reig, L. (2009). Sequential egocentric strategy is acquired as early as allocentric strategy: Parallel acquisition of these two navigation strategies. *Hippocampus*, 19, 1199–1211. <https://doi.org/10.1002/hipo.20595>.
- Jacobs, W. J., Laurance, H. E., & Thomas, K. G. F. (1997). Place learning in virtual space I: Acquisition, overshadowing, and transfer. *Learning & Motivation*, 28, 521–541.
- Jordan, K., Schadow, J., Wuestenberg, T., Heinze, H. J., & Jancke, L. (2004). Different cortical activations for subjects using allocentric or egocentric strategies in a virtual navigation task. *NeuroReport*, 15(1), 135–140.
- Kamin, L. J. (1968). Attention-like processes in classical conditioning. *Miami Symposium on the Prediction of Behavior: Aversive Stimulation*, 9–31.
- Kamin, L. J. (1969). Predictability, surprise, attention and conditioning. In B. A. Campbell, & R. M. Church (Eds.), *Punishment and aversive behavior* (pp. 279–296). New York: Appleton-Century-Crofts.
- Karnath, H. O., Wallesch, C. W., & Zimmermann, P. (1991). Mental planning and anticipatory processes with acute and chronic frontal lobe lesions: A comparison of maze performance in routine and non-routine situations. *Neuropsychologia*, 29(4), 271–290. [https://doi.org/10.1016/0028-3932\(91\)90042-7](https://doi.org/10.1016/0028-3932(91)90042-7).
- Klatzky, R. L. (1998). Allocentric and egocentric spatial representations: definitions, distinctions, and interconnections. *Spatial cognition* (pp. 1–17). Berlin, Heidelberg: Springer.
- Kolb, B., Sutherland, R. J., & Whishaw, I. Q. (1983). A comparison of the contributions of the frontal and parietal association cortex to spatial localization in rats. *Behavioral Neuroscience*, 97(1), 13–27. <https://doi.org/10.1037/0735-7044.97.1.13>.
- Korthauer, L., Nowak, N., Frahm, M., & Driscoll, I. (2017). Cognitive correlates of spatial navigation: Associations between executive functioning and the virtual Morris Water Task. *Behavioural Brain Research*, 317, 470–478. <https://doi.org/10.1016/j.bbr.2016.10.007>.
- Krigolson, O. E., Hassall, C. D., & Handy, T. C. (2014). How we learn to make decisions: Rapid propagation of reinforcement learning prediction errors in humans. *Journal of Cognitive Neuroscience*, 26(3), 635–644. https://doi.org/10.1162/jocn_a.00509.
- Loftus, G. R., & Masson, M. E. J. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review*, 1(4), 476–490.
- Luck, S. J. (2014). *An introduction to the event-related potential technique* (2nd ed.). MIT Press.
- Maaswinkel, H., & Whishaw, I. Q. (1999). Homing with locale, taxon, and dead reckoning strategies by foraging rats: Sensory hierarchy in spatial navigation. *Behavioural Brain Research*, 99, 143–152.
- McClelland, J. L., McNaughton, B. L., & O’Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, 102(3), 419–457.
- McDonald, R. J., & White, N. M. (1994). Parallel information processing in the water maze: evidence for independent memory systems involving dorsal striatum and hippocampus. *Behavioral and Neural Biology*, 61(3), 260–270. [https://doi.org/10.1016/S0163-1047\(05\)80009-3](https://doi.org/10.1016/S0163-1047(05)80009-3).
- Miltner, W. H. R., Braun, C. H., & Coles, M. G. H. (1997). Event-related brain potentials following incorrect feedback in a time-estimation task: Evidence for a ‘Generic’ neural system for error detection. *Journal of Cognitive Neuroscience*, 9(6), 788–798.
- Morris, R. (1984). Developments of a water-maze procedure for studying spatial learning in the rat. *Journal of Neuroscience Methods*, 11, 47–60.
- O’Keefe, J., & Nadel, L. (1978). *The hippocampus as a cognitive map*. Oxford: Clarendon Press.
- Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, 2011, 156869. <https://doi.org/10.1155/2011/156869>.
- Packard, M. G., & McGaugh, J. L. (1996). Inactivation of hippocampus or caudate nucleus with lidocaine differentially affects expression of place and response learning. *Neurobiology of Learning and Memory*, 65, 65–72.
- Pearce, J. M. (2009). The 36th sir frederick bartlett lecture: An associative analysis of spatial learning. *Quarterly Journal of Experimental Psychology*, 62(9), 1665–1684. <https://doi.org/10.1080/17470210902805589>.
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spatial Vision*, 10(4), 437–442.
- Proudfit, G. H. (2015). The reward positivity: From basic research on reward to a biomarker for depression. *Psychophysiology*, 52(4), 449–459. <https://doi.org/10.1111/psyp.12370>.
- Redhead, E. S., & Hamilton, D. A. (2007). Interaction between locale and taxon strategies in human spatial learning. *Learning and Motivation*, 38(3), 262–283. <https://doi.org/10.1016/j.lmot.2006.11.003>.
- Rescorla, R. A., & Wagner, A. R. (1972). Theory of pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black, & W. F. Prokasy (Eds.), *Classical conditioning II: current research and theory* (pp. 64–99). New York: Appleton-Century-Crofts.
- Roberts, A. D. L., & Pearce, J. M. (1999). Blocking in the morris swimming pool. *Journal of Experimental Psychology Animal Behavior Processes*, 25(2), 225–235. <https://doi.org/10.1037/0097-7403.25.2.225>.
- Schmitzer-Torbert, N. (2007). Place and response learning in human virtual navigation: Behavioral measures and gender differences. *Behavioural Neuroscience*, 121(2), 277–290. <https://doi.org/10.1037/0735-7044.121.2.277>.
- Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science (New York, N.Y.)*, 275(5306), 1593–1599. <https://doi.org/10.1126/SCIENCE.275.5306.1593>.

- Spencer, K., Dien, J., & Donchin, E. (2001). Spatiotemporal analysis of the late ERP to deviant stimuli. *Psychophysiology*, *38*, 343–358. <https://doi.org/10.1111/1469-8986.3820343>.
- Spiers, H. J., & Maguire, E. A. (2006). Thoughts, behaviour, and brain dynamics during navigation in the real world. *NeuroImage*, *31*(4), 1826–1840. <https://doi.org/10.1016/j.neuroimage.2006.01.037>.
- Spriggs, M. J., Kirk, I. J., & Skelton, R. W. (2018). Hex Maze: a new virtual maze able to track acquisition and usage of three navigation strategies. *Behavioural Brain Research*, *339*, 195–206. <https://doi.org/10.1016/j.bbr.2017.11.041>.
- Sutherland, R. J., Whishaw, I. Q., & Kolb, B. (1988). Contributions of cingulate cortex to two forms of spatial learning and memory. *The Journal of Neuroscience*, *8*(8), 1863–1872.
- Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: an introduction*. MIT Press.
- Tolman, E. C. (1948). Cognitive maps in rats and men. *The Psychological Review*, *55*(4), 189–208.
- Torok, A., Obor, A. K., Persa, G., Galambos, P., Baranyi, P., Csepe, V., & Honbolygo, F. (2017). Temporal dynamics of object location processing in allocentric reference frame. *Psychophysiology*, *54*, 1346–1358. <https://doi.org/10.1111/psyp.12886>.
- Trullier, O., Wiener, S. I., Berthoz, A., & Meyer, J.-A. (1997). Biologically-based artificial navigation systems: Review and prospects. *Progress in Neurobiology*, *51*(5), 483–544.
- Umemoto, A., HajiHosseini, A., Yates, M. E., & Holroyd, C. B. (2017). Reward-based contextual learning supported by anterior cingulate cortex. *Cognitive, Affective and Behavioral Neuroscience*, *17*(3), 642–651. <https://doi.org/10.3758/s13415-017-0502-3>.
- Vassena, E., Holroyd, C. B., & Alexander, W. H. (2017). Computational models of anterior cingulate cortex: At the crossroads between prediction and effort. *Frontiers in Neuroscience*, *11*, 316. <https://doi.org/10.3389/fnins.2017.00316>.
- Vogt, B. A. (2016). Midcingulate cortex: structure, connections, homologues, functions and diseases. *Journal of Chemical Neuroanatomy*, *74*, 28–46. <https://doi.org/10.1016/J.JCHEMNEU.2016.01.010>.
- Wang, R. F. (2004). Between reality and imagination: When is spatial updating automatic? *Perception and Psychophysics*, *66*(1), 68–76. <https://doi.org/10.3758/BF03194862>.
- Wang, R. F., & Brockmole, J. R. (2003). Human navigation in nested environments. *Journal of Experimental Psychology: Learning Memory and Cognition*, *29*(3), 398–404. <https://doi.org/10.1037/0278-7393.29.3.398>.