Modeling Mental Navigation in Scenes with Multiple Objects

Patrick Byrne
byrne@psychology.mcmaster.ca

Suzanna Becker
becker@mcmaster.ca

Department of Psychology, McMaster University, Hamilton, Ontario, Canada

Various lines of evidence indicate that animals process spatial information regarding object locations differently from spatial information regarding environmental boundaries or landmarks. Following Wang and Spelke’s (2002) observation that spatial updating of egocentric representations appears to lie at the heart of many navigational tasks in many species, including humans, we postulate a neural circuit that can support this computation in parietal cortex, assuming that egocentric representations of multiple objects can be maintained in prefrontal cortex in spatial working memory (not simulated here). Our method is a generalization of an earlier model by Droulez and Berthoz (1991), with extensions to support observer rotation. We can thereby simulate perspective transformation of working memory representations of object coordinates based on an egomotion signal presumed to be generated via mental navigation. This biologically plausible transformation would allow a subject to recall the locations of previously viewed objects from novel viewpoints reached via imagined, discontinuous, or disoriented displacement. Finally, we discuss how this model can account for a wide range of experimental findings regarding memory for object locations, and we present several predictions made by the model.

1 Introduction

Spatial reasoning is of paramount importance in nearly all aspects of human behavior, from planning and navigating a complex route through some environment in order to reach a relevant goal, to simply grasping a nearby object. The set of all entities that comprise our spatial surroundings may be divided into subsets in arbitrarily many ways. For the present purposes, we will partition these entities into environmental boundaries/landmarks and objects. There is substantial empirical evidence as to how the brain represents and processes the former (for a review, see Burgess, Becker, King, & O’Keefe, 2001). For example, O’Keefe & Dostrovsky (1971) found neurons in the hippocampus of the rat that respond to the rat’s location in space. O’Keefe and Nadel (1978) argue that this collection of “place cells” forms a
cognitive map and is the rat’s internal allocentric map of the environment. Evidence of view-invariant hippocampal place cells has also been found in nonhuman primates (Ono, Nakamura, Nishijo, & Eifuku, 1993) and in human hippocampus (Ekstrom et al., 2003).

For the case of short-term object location representation, the focus of this article, empirical evidence strongly indicates involvement of the posterior parietal cortex. For example, Sabes, Breznen, and Andersen (2002) perform single unit recordings that demonstrate that area LIP of monkey cortex encodes saccade targets in retinotopic coordinates. More generally, Colby and Goldberg (1999) review evidence showing that object locations are represented in a variety of reference frames in parietal cortex, while Andersen, Shenoy, Snyder, Bradley, and Crowell (1999) review evidence suggesting that area 7a incorporates vestibular information to maintain localization of visual stimuli in a world-centered reference frame. Finally, Goodale and Milner (1992) review data suggesting that the dorsal stream from striate to posterior parietal cortex is responsible for the relatively short timescale sensorimotor transformations used while performing visually guided actions directed at objects in the environment.

In this article, we first review empirical evidence from neuropsychological and electrophysiological studies as to the nature and locus of object representations in the brain. We then present a biologically plausible computational model that allows for the transformation of object coordinates maintained in spatial working memory (WM). The function of this transformation is to update WM representations of object coordinates in egocentric space while the subject mentally navigates through the environment. This transformation is driven by the products of mental navigation (presumably some mentally generated equivalent to vestibular, proprioceptive, and motor efference information). Next, we perform simulations demonstrating the functioning of the model and how error is introduced into object coordinate representations by the model. Finally, we discuss how the model can account for various experimental findings.

The representation of object location information in the brain appears, at least under certain circumstances, to be quite different from the representation of environmental boundary information. A series of experiments performed by Wang & Spelke (2000) provides insight into how humans differentially process object location and environmental boundary information. For each experiment, the authors placed a few objects around a small room. Subjects were allowed to explore the room and studied objects’ locations for as long as they pleased. They were then brought to the center of the room, blindfolded, rotated a small amount, and asked to point to the various objects as they were called out in a random order by the experimenters. Following this, subjects were required to sit in a swivel chair fixed at the center of the room and disorient themselves via a 1 minute self-induced rotation. They were again asked to point to the objects in a predetermined random order. In addition to objects, subjects were also required to point to
room corners in some experiments. The main findings can be summarized as follows: After the initial rotation (without disorientation), subjects could point to objects and room corners with relatively high accuracy, implying that they had encoded correctly the information that they were asked to. After disorientation, subjects could no longer accurately point to the location of objects or room corners. However, analyses of the consistency of the pointing errors indicated that the relative configuration of the room corners could be accurately recalled (for nonrectangular as well as rectangular rooms), whereas the relative object configuration could not be. This supports a configural representation of geometric feature locations and nonconfigural or independent representations of object locations. Furthermore, when subjects were reoriented with a bright light (still blindfolded), they could accurately recall the absolute and relative locations of room corners but neither the relative nor absolute locations of objects. When the bright light was left on throughout the disorientation procedure, subjects could recall the relative and absolute locations of both objects and room corners.

The results of Wang and Spelke (2000) experiments indicate that subjects could rapidly form an accurate internal representation of environmental boundaries and object locations. However, it appears that during mental transformations of spatial information in trials where the subject is not continuously oriented, information about a given object location is updated independent of information regarding other objects and of the environment. In this way, each object location is subject to independent transformation error. This is in contrast to environmental geometric cues such as room corners, which appear to be updated as a coherent whole.

Additional evidence that object location information is handled differently from environmental boundary information comes from an experiment by Shelton and McNamara (2001). Subjects were brought into a room with various objects placed at different locations inside and allowed to observe the objects from various predetermined viewpoints. After leaving the room, they were asked to imagine standing at a given object facing a second object and to point to where a third object would be relative to themselves. Subjects performed best when their imagined viewpoint was aligned with the original viewpoint from which they observed the object configuration. This suggests the possibility that subjects are storing the object configuration in head-centered egocentric coordinates and that they must transform it (introducing error) when they imagine observing it from other viewpoints. We say head centered here because subjects were asked to imagine facing an object, which implies pointing their head toward it; they were not asked to constrain their gaze direction, so neither retinal nor body-centered coordinates are implied. Head-centered coordinates are also implied by the unit recordings of Funahasi, Bruce, and Goldman-Rakic (1989), as we mention in the next section. Another possible interpretation of Shelton and McNamara’s (2001) results is that subjects form a view-based snapshot of the entire presentation scene, which can be used for later matching. Wang and Spelke
1854 P. Byrne and S. Becker

(2002) review evidence that humans do seem to make use of such snapshots, at least under certain circumstances. However, we do not investigate this issue any further here.

In order to build a computational model of mental navigation that explains empirical results such as those of Wang and Spelke (2000) and Shelton and McNamara (2001), we first require a more complete understanding of how the brain represents objects. We next review empirical evidence as to the nature and locus of object representations in the brain.

2 Objects and Spatial Working Memory

A number of experiments suggest a transition from short-term spatial WM to long-term memory representations of objects after several minutes of study time (Smith & Milner, 1989; Crane, Milner, & Leonard, 1995; Bohbot et al., 1998). The evidence provided by these studies suggests that medial temporal lobe structures are essential to location memory over periods of several minutes (≈ 4 minutes) or greater but are less relevant over shorter timescales.

To investigate the nature of short-term object location memory, we consider the evidence from functional imaging and unit recording studies. In an fMRI experiment performed by Galati et al. (2000), subjects were required to report the location of a vertical bar flashed before them for 150 ms relative to their midsagittal plane. During this task, several frontal and parietal regions were more active on the location task, relative to a control color decision involving the same stimuli. Furthermore, Sala, Ràmà, and Courtney (2003) presented subjects with a sequence of flashes and asked them to recall the location or identity of what was shown (picture of a house or a face) three flashes back. During location recall, fMRI scans revealed activation in the superior portion of the intraparietal sulcus (IPS) and in the superior frontal sulcus, as well as other areas. During identity recall, activation was found in the inferior and medial frontal gyrus, as well as other areas. These observations indicate that areas of frontal and parietal cortices are of key importance in generating and maintaining internal representations of spatial locations, at least for short periods of time. The role of frontal cortical areas in this process is reviewed by Levy and Goldman-Rakic (2000), who argue that the principal sulcus (area 46) plays a crucial role in spatial WM and that Walker’s areas 12 and 45 (the inferior convexity) play a crucial role in object identity WM. In particular, a unit recording study by Funahasi et al. (1989) showed that neurons in the principal sulcus of the monkey appear to code for egocentric spatial locations in head-centered coordinates. Furthermore, a human study by Oliveri et al. (2001), requiring subjects to remember the position of a flash two steps back in a sequence, found that only when transcranial magnetic stimulation (TMS) was applied to the dorsolateral prefrontal cortex (DLPFC) was accuracy affected, although TMS applied to several different brain regions affected reaction times. These re-
sults suggest that egocentric representations of spatial locations are maintained in DLPFC. Unfortunately, the role of parietal cortical areas in spatial WM is somewhat less clear. It is known, however, that neurons in areas VIP and LIP of the IPS show receptive fields in head-centered coordinates (see Burgess, Jeffery, and O’Keefe, 1999, for an overview). Also, Chaffe and Goldman-Rakic (1998) showed that when monkeys are required to hold a target location in memory for a short delay before making a saccade to it, neurons in area 8a (near the principle sulcus of prefrontal cortex) and in area 7ip (near the IPS) become active and show temporally varying activation levels over the delay period. Neurons in both regions show spatial selectivity similar to that found in the principal sulcus of monkeys by Funahasi et al. (1989).

From the above evidence, it seems plausible that object locations are initially represented in parietal cortex, and if sufficient attention is allocated to them, then their locations and identities are maintained in WM in DLPFC. In particular, their locations are represented egocentrically (as Shelton and McNamara’s, 2001, work might indicate) in one area of DLPFC, while maintenance of their identities involves a different area of DLPFC consistent with Goldman-Rakic’s hypothesis that the ventral/dorsal (what/where) distinction persists into the DLPFC. In order to mentally manipulate the positions of objects stored in WM, it seems reasonable to assume that a circuit involving areas of the parietal cortex (especially the IPS) would be involved, given this area’s involvement in spatial WM and its known ability to represent locations in multiple reference frames. In particular, we hypothesize that object coordinates are maintained in WM egocentrically in DLPFC and manipulated and represented more transiently in the vicinity of the IPS.

3 Model

In this work, we develop a computational account of memory for object locations in the face of certain types of viewer motion, specifically those in which the motion is imagined or those in which the subject does not remain continuously oriented throughout. At a minimum, the model should provide an explanation of Shelton and McNamara’s (2001) finding that subjects more accurately recall object positions when asked to do so from a viewpoint in which they previously observed the object configuration, and of Wang and Spelke’s (2000) finding regarding the pattern of errors made by disoriented subjects recalling object configurations. One possible way in which subjects might perform either the Shelton and McNamara or the Wang and Spelke task is by mental navigation. In contrast to a simple mental rotation of the object array, mental navigation involves imagined egomotion. This entails making mental viewpoint transformations while simultaneously updating WM representations of egocentric object coordinates. More specifically, subjects could make a WM snapshot of object locations from a given viewpoint, and when asked to recall object locations from a new viewpoint, they could
mentally navigate from the initial viewpoint to the new viewpoint and use the same motion signal driving mental navigation to simultaneously drive a transformation of egocentric object coordinates. In the case of the Wang and Spelke tasks, the new viewpoint would have to be estimated due to the disorientation procedure. Also, assuming a serial updating procedure in which the transformation just described must be repeated for each individual object allows for a possible explanation of the finding that object configurations could not be accurately recalled after disorientation and that recall errors were not systematic, indicating lack of a configurational representation. This will be discussed in more detail in section 5.

Our hypothesis requires neural circuits that can represent the environment allocentrically, to allow for mental navigation through the environment. Additional circuits are required to maintain egocentric object coordinates and transform these coordinates, one object at a time, based on the egomotion signal that drives mental navigation. To this end, models of real navigation based on internal allocentric cognitive maps have been developed for rats and humans (see Voicu, 2003, and Burgess, Donnett, & O’Keefe, 1997, for examples). Given the evidence that both real and imagined spatial tasks invoke nearly the same cortical circuitry (see, e.g., Stippich, Ochmann, & Sartor, 2002; Ino et al., 2002; Mellet et al., 2000; Kreiman, Koch, & Fried, 2000), these models of real navigation will be assumed to be applicable to mental navigation. We now require a neural circuit that performs the object location transformation.

To begin, we must decide how egocentric object coordinates are to be represented. Thelen, Schöner, Scheier, and Smith (2001) and Compte, Brunel, Goldman-Rakic, and Wang (2000) have created models of spatial WM that hold location as a bump of activity in a topographically organized neural circuit, in which each neuron represents a location or direction in egocentric space. Becker and Burgess (2001) model the parietal egocentric map in this way, and similarly the organism’s location in allocentric space is represented by a Gaussian bump of activity over an array of hippocampal place cells. Such bump attractor networks have been used by others in models of hippocampus as well (Zhang, 1996; Samsonovich & McNaughton, 1997). Here also it will be assumed that the model contains a main layer of neurons, each of which will be preassigned a unique location on a Cartesian grid covering head-centered egocentric space. We will represent object location in egocentric space as a Gaussian bump of activity in the main neuron layer. By definition, the person’s coordinates in this map are the origin, and their orientation will be taken as facing along the positive y-axis. The existence of these main-layer neurons in parietal cortex is supported by the electrophysiological recordings of Chaffee and Goldman-Rakic (1998).

In addition to a representation of object locations, our model will be adapted from a model by Droulez and Berthoz (1991), which could easily be implemented to handle translational movements of the observer but requires a nontrivial extension to handle rotational motion. To derive the
model, first consider an observer standing at the origin, $O$, of some reference frame and facing along the positive $y$-axis (see Figure 1). The position of an object at a point, $P$, will be denoted by the vector, $r_P$. If the observer moves by an amount $r_T$ and rotates by an amount $\theta$, then we will call the new egocentric reference frame the primed frame. The position of the object with regard to this new frame is given by

$$r'_P = r_P - r_T.$$  \hfill (3.1)

Reexpressed in terms of Cartesian $x$, $y$ components, this is

$$x'_p i' + y'_p j' = (x_p - x_T) i + (y_p - y_T) j.$$  \hfill (3.2)

where $i$ and $j$ are basis vectors oriented along the $x$- and $y$-axes, respectively. Making an appropriate change of basis on the right-hand side yields

$$x'_p = (x_p - x_T) \cos \theta + (y_p - y_T) \sin \theta$$

$$y'_p = (y_p - y_T) \cos \theta - (x_p - x_T) \sin \theta.$$  \hfill (3.3)

Next, we assume that the observer’s viewpoint shift occurs over a time interval $\Delta t \ll 1$, so that they are standing at $O$ facing in the $j$ direction at time $t$ and at $O'$ facing in the $j'$ direction at time $t + \Delta t$. Assuming that
egocentric velocities vary smoothly enough so that for any time, $t' \in [t, t + \Delta t]$, we can write
\begin{align*}
v_x(t') &= v_x^0 + f_x(t'), \\
v_y(t') &= v_y^0 + f_y(t'), \\
\omega(t') &= \omega^0 + f\omega(t'),
\end{align*}
(3.4)
where $v_x(t')$, $v_y(t')$, and $\omega(t')$ are the translational and rotational velocities of the observer as measured in the $O$ frame, and the $f$ functions vary by $O(\Delta t)$ over the interval and satisfy $f(t' = t) = 0$; then we have
\begin{align*}
x'(p) &= x(p) - \frac{\left[ v_x^0 - \omega^0 y \right]}{\Delta t} \Delta t + O\left( \frac{\Delta t^2}{\Delta t} \right) \\
y'(p) &= y(p) - \frac{\left[ v_y^0 + \omega^0 x \right]}{\Delta t} \Delta t + O\left( \frac{\Delta t^2}{\Delta t} \right).
\end{align*}
(3.5)
Of course, the primed position variables are simply the egocentric object coordinates at time $t + \Delta t$, the unprimed position variables are simply the egocentric object coordinates at time $t$, and the velocity variables are the egocentric velocities at time $t$. Therefore, we can rewrite equations 3.5 as
\begin{align*}
x(t + \Delta t) &= x(t) - \left[ v_x(t) - \omega(t)y(t) \right] \Delta t + O\left( \Delta t^2 \right) \\
y(t + \Delta t) &= y(t) - \left[ v_y(t) + \omega(t)x(t) \right] \Delta t + O\left( \Delta t^2 \right),
\end{align*}
(3.6)
where $x(t)$ and $y(t)$ refer to the egocentric object coordinates at time $t$ and $\{v_x(t), v_y(t), \omega(t)\}$ are the egocentrically measured velocities at time $t$.

We are interested in representing the egocentric location of an object by a bump of activity across a population of neurons moving around within the observer’s egocentric reference frame over time. This activity can be expressed as a function $A(x, y, t)$ over egocentric coordinates and time. Given the value of $A(x, y, t)$, the value of $A(x, y, t + \Delta t)$ can easily be found by applying the inverse of equations 3.6 to the arguments of $A$, as follows:
\begin{align*}
A(x, y, t + \Delta t) &= \\
A(x + [v_x(t) - \omega(t)y] \Delta t + O(\Delta t^2), \\
y + [v_y(t) + \omega(t)x] \Delta t + O(\Delta t^2), t).
\end{align*}
(3.7)
The right-hand-side of this equation may be expanded using the second mean value theorem. Such an expansion will be valid as long as the second derivatives of $A$ and the coefficients of $\Delta t$ remain small. Notice, however, that these coefficients actually depend on location, so for distant objects,
more error will be introduced into the approximation. We are thus left with

\[ A(x, y, t + \Delta t) = A(x, y, t) + \left[ v_x(t) - \omega(t) y \right] \Delta t \frac{\partial A}{\partial x}(x, y, t) + \left[ v_y(t) + \omega(t) x \right] \Delta t \frac{\partial A}{\partial y}(x, y, t) + O(\Delta t^2). \] (3.8)

Recall that space in this model is to be represented as a Cartesian grid of points, with a single main layer neuron allocated to each point. In an attempt to follow the notation of Droulez and Berthoz (1991), we note that each of these positions and, hence, the corresponding main layer neurons can be labeled with a positive integer, starting at 1 for the position (neuron) with the smallest \( x \) and \( y \) value and increasing with position in the direction of increasing \( x \) until one row is complete. The numbering continues on the row with the next lowest \( y \) value until the last position is reached (see Figure 2). By doing this, equation 3.8 can be discretized as follows,

\[ A(x_i, y_i, t + \Delta t) \approx \sum_j \left( a_{ij} + b_{ij} v_x(t) + c_{ij} v_y(t) + d_{ij} \omega(t) \right) A(x_i + dx_{ij}, y_i + dy_{ij}, t), \] (3.9)

where \( x_i \) and \( y_i \) are the \( x \) and \( y \) coordinates of the \( i \)th location and \( dx_{ij} \) and \( dy_{ij} \) are the distances from the \( i \)th location to the \( j \)th location along the \( x \) and \( y \) directions. The values of the \( a, b, c, \) and \( d \) coefficients are determined by the approximation used to calculate the gradients of \( A \). For example, to approximate equation 3.8 using a centered difference rule for the partial derivatives, the coefficients should be selected as

\[ a_{ij} = \delta_{ij}, \]
\[ b_{ij} = \frac{\Delta t}{2dx_{ij}} (\delta_{i+1,j} - \delta_{i-1,j}), \]
\[ c_{ij} = \frac{\Delta t}{2dy_{ij}} (\delta_{i+N_x,j} - \delta_{i-N_x,j}), \]
\[ d_{ij} = \frac{\Delta t}{2} \left[ \frac{x_i}{dy_{ij}} (\delta_{i+1,j} - \delta_{i-1,j}) - \frac{y_i}{dx_{ij}} (\delta_{i+1,j} - \delta_{i-1,j}) \right] \] (3.10)

where \( \delta_{ij} \) is the Kronecker delta function and \( N_x \) is the number of neurons spanning the \( x \)-direction.

Notice that equation 3.9 can be rewritten as
Figure 2: Schematic of proposed neural network. The main layer neurons are interconnected with weights “a”. They are also each connected to the neurons with the same x and y coordinates in the other layers via unit weights. Neurons in the translational velocity layer multiply current egocentric velocity with the corresponding main neuron activation from the current time step and feedback with weights “b”. Vy neurons have been left out of this diagram for clarity. Finally, neurons in the angular velocity layer multiply main layer activations by current head rotation speed and feedback with weights “c”.

\[ A^M(x_i, y_i, t + \Delta t) = \sum_j a_{ij} A^M(x_i + dx_{ij}, y_i + dy_{ij}, t) \]
\[ + b_{ij} A^V(x_i + dx_{ij}, y_i + dy_{ij}, t) \]
\[ + c_{ij} A^V(x_i + dx_{ij}, y_i + dy_{ij}, t) \]
\[ + d_{ij} A^\omega(x_i + dx_{ij}, y_i + dy_{ij}, t) \]
\[ A^V(x_i, y_i, t) = v_x(t) A^M(x_i, y_i, t) \]
\[ A^V(x_i, y_i, t) = v_y(t) A^M(x_i, y_i, t) \]
\[ A^\omega(x_i, y_i, t) = \omega(t) A^M(x_i, y_i, t), \] (3.11)
where we have relabeled $A \rightarrow A^M$. These equations can now be viewed as a neural network with four layers, and the parameters $a, b, c,$ and $d$ can be viewed as weights (see Figure 2 for a schematic), each of which couples a spatial location at the input processing layer with another spatial location in the transformation layer. Notice, however, that these weights are modulated by a scalar velocity value (i.e., the network contains neurons that perform multiplicative operations on their inputs). Neurons with this ability have been postulated in models of visual cortex (Mel, 1993, 1994). It is also thought that neurons that show such gain modulation are important in performing coordinate transformations in parietal cortex (Salinas & Abbott, 1996; Andersen, Essick, & Siegel, 1985). Additionally, we are assuming that the signal used to drive the transformation is an egocentric velocity signal. Although the velocity signal used to drive mental navigation in our model is not vestibular in nature but rather mentally generated, it seems reasonable to make this assumption because vestibular information is known to project to Brodman’s area 7, which borders portions of the IPS (Kawano, Sasaki, & Yamashita, 1980, 1984). If neural circuitry is adapted to using such information in this area, then it seems reasonable that a mentally generated equivalent could also readily be made use of. Note that although we represent velocity as three scalar values, along the $x, y,$ and $\theta$ directions, the model could easily be extended to accommodate coarse coding of the velocity signal. Finally, we are assuming a “Cartesian” distribution of neurons over egocentric space. It seems unlikely that cortex is organized in such a way; however, this organization is convenient for the present purposes, and we do not expect the behavior of our model to be qualitatively changed by the specific choice of neural distribution so long as each region of egocentric space is represented by a sufficiently high neural density. Quantitatively, for a different neural-spatial map, we would expect more transformation error to be introduced in areas where space is more sparsely represented.

Satisfactory performance of our model can be attained using the weights defined by equation 3.10. However, further improvements may be made by applying a gradient descent algorithm to refine the connection strengths. To be specific, at each iteration of the algorithm, a gaussian bump of activity is injected into the main layer at a random location, random velocities are applied to the network, the activity is updated for one time step, and the resulting main layer activation is compared to an exact calculation of what it should be. This comparison is used to calculate the value of an error function given by

$$E (\{a_{ij}, b_{ij}, c_{ij}, d_{ij}\}) = \sum_i \left[ A^M (x_i, y_i) - A^M (x_i, y_i) \right]^2,$$

(3.12)

where $A^M$ is the target activation determined as described below. The gradient of this error function with regard to weights can be calculated analyt-
ically, and the weight vector can thus be updated to reduce the value of the error.

To determine the target function, \( A^M \), we note that it too must be a gaussian with the same width as the input gaussian, but that it must be centered at a slightly shifted location. This location can be determined by dividing equations 3.6 by \( \Delta t \) and taking the limit as \( \Delta t \to 0 \). The result is a set of two exact differential equations for the egocentric coordinates of the peak of the gaussian bump in time. The constant velocity solutions of these equations are

\[
\begin{align*}
    x(t) &= \left( x(0) + \frac{\dot{y}}{\omega} \right) \cos(\omega t) \,+ \left( y(0) - \frac{\dot{x}}{\omega} \right) \sin(\omega t) - \frac{\dot{y}}{\omega}, \\
    y(t) &= \left( y(0) - \frac{\dot{x}}{\omega} \right) \cos(\omega t) \,- \left( x(0) + \frac{\dot{y}}{\omega} \right) \sin(\omega t) + \frac{\dot{x}}{\omega}.
\end{align*}
\] (3.13)

As one final point regarding the training method, it should be noted that the gradient descent method used here can be made equivalent to a Hebbian-like mechanism by the addition of an extra layer of “input” neurons that compare main layer activity to an external target signal (see Droulez and Berthoz, 1991).

4 Simulations

In the simulations reported here, a network with 961 main layer neurons arranged in a 31 \times 31 lattice in which each neuron represents a spatial area of one unit\(^2\) was trained using the method described in the previous section. Each of the three transformation layers also contained 31 \times 31 neurons. The width of each gaussian activation bump injected into the main layer for training was chosen randomly from the interval \([2, 4]\) units (widths below 2 units result in large error due to the discrete nature of the lattice; large widths are not useful in representing localized objects), with the peak height always set to unity. Angular velocities were chosen randomly from the interval \([-1, 1]\) radians per second, while translational velocities were chosen from the interval \([-3, 3]\) units per second. These ranges were chosen to be reasonable for a human subject in the case of 1 unit = 1 meter, but the exact values, which have little effect on training results, were chosen arbitrarily. A time step of 10 ms was used for training and for all simulations. With initial weights set to the values defined in equations 3.10, the simple gradient descent algorithm used here was found to converge after approximately 500,000 random infinitesimal transformations when the learning rate was set to \( 5 \times 10^{-4} \). Although error, as defined by equation 3.12, was reduced by a factor of about 4 during this procedure, all weights remained within 4\% of their initial values after training was complete.
As a demonstration of the functioning of the model, a simulation was carried out in which the egocentric location of a stationary object was maintained while the subject performed a sequence of imagined motions in space (see Figure 3). To begin, an object was placed 5 units away from the subject, directly to his or her right. The subject turned 90 degrees to the left, moved forward by 5 units, and then turned 90 degrees to the right. The predicted trajectory of the stationary object through egocentric space, as computed by the model, can be seen in Figure 4, where main layer activations, sampled every 20 time steps, have been superimposed. Initial and final activations are shown in Figures 5 and 6.

From Figure 6, we see that the bump of activity representing the object’s egocentric coordinates is somewhat attenuated and deformed by the transformation procedure. In fact, error is introduced in the simulation by two distinct mechanisms. First, the truncated leading-order terms in equation 3.6 contain factors of $v^2_x$, $v^2_y$, and $\omega^2$, so simulation error should increase quadratically in velocity for large velocities (as long as velocity $\times \Delta t < 1$, after which point error will increase more rapidly with increasing velocity). Second, at low velocities, truncating the $O(\Delta t^2)$ terms will have little effect in terms of introducing error compared with the effect of approximating...
Figure 4: Main layer activation representing the egocentric location of a stationary object, from the perspective of the moving observer depicted in Figure 3. Main layer neurons are sampled every 20 time steps and superimposed. The numbered arrows correspond to the numbered observer motions in Figure 3.

Figure 5: Initial main layer activation before navigation begins.
the first derivatives of activation using information from a discrete lattice of neurons. In this case, error should increase linearly with the number of time steps required for a given transformation or, equivalently, as the inverse of velocity. A number of simulations were performed in order to investigate these effects. The results of two of these simulations will now be discussed. For the first, an object was placed 5 units to the egocentric right of an observer who then turned 90 degrees to the left. For the second, an object was placed 5 units to the egocentric left of an observer who then moved to the egocentric right, a distance that is equal to the arc length traveled by the object in the first simulation. In both cases, the object was represented by a gaussian activation bump with a width of 2 units. Squared error is calculated for the final activation using equation 3.12 and plotted against velocity in Figures 7 and 8. In both cases, the optimum velocity is around 2 units per second. The optimum velocity actually remains in the [1.5, 3.5] units per second interval for a wide variety of paths. Finally, the solid curves in Figures 7 and 8, which are of the form

\[ f(x) = \frac{a_1}{x} + a_2 (x - a_3)^2 + a_4, \]  

(4.1)

were fit to the data in order to demonstrate that transformation error depends on velocity in the way described above.

So far we have discussed the functioning of our model in isolation. In the brain, we expect that object location is maintained in DLPFC, as we have discussed in previous sections. One possible way in which this information
is maintained is as a collection of activity bumps over continuous attractor neural networks residing in the DLPFC, although we do not speculate on the details here. If the recurrent main layer weights ("a" weights) were reduced in strength, then an activity bump in our model would require external driving to be maintained. This driving could come directly from the DLPFC activity. Via feedback connections, the movement of the activity bump in our circuit could cause the DLPFC bump representation of object location to move along with it. This scenario does not necessarily imply that the representation of the original egocentric object coordinates is lost. Pickering, Gathercole, Hall, and Lloyd (2001) perform experiments and review evidence that they claim implies that visuo-spatial working memory is dissociated into static and dynamic components. For our case, the updating of object locations might occur in something analogous to the dynamic component, while a static snapshot might be maintained elsewhere. Simulations of such DLPFC-parietal interactions are currently being implemented in our lab.

Figure 7: Transformation error produced by the network for an observer rotation of 90 degrees. Triangles are simulation data, and the solid curve is the fit to equation 4.1.
5 Discussion

In this article, we have postulated a biologically plausible computation, implemented as a neural circuit, that explains how the contents of spatial WM can be updated to keep track of the egocentric coordinates of remembered locations during mental navigation over timescales shorter than 4 minutes. Error can be introduced into the representation of object location coordinates during transformation by a number of different mechanisms. We have demonstrated that the discrete nature of the representation of space and the exclusion of higher-order velocity information in a rate-coded neural circuit of this kind leads to a nonzero error regardless of the velocity of the subject through space. This error appears to affect peak height predominantly, but the effect on final peak location is relatively small in the simulations performed here. There is, however, an optimum velocity range for performing transformations. Other potential sources of error are internal and external noise, which we have not studied here.

A final source of error in the kind of object location transformation postulated here is the potential error introduced by the mental navigation pro-
procedure itself. This process must depend on an internal allocentric map of
the subject’s environment. Evidence indicates that this map contains inac-
curacies and imprecisions (Hirtle & Jonides, 1985), perhaps more so in novel
environments (Hartley, Burgess, Lever, Carucci, & O’Keefe, 2000). Presum-
ably, tasks requiring precise mental navigation would also suffer from these
errors.

Shelton and McNamara’s (2001) finding that subjects more accurately re-
member object configurations when tested from viewpoints in which they
have actually observed the object configuration is readily explained by our
model. Subjects would simply store an egocentric view of the studied pre-
sentation in WM and then mentally navigate to the test viewpoint while
updating object locations using the proposed parietal circuit. Note that this
circuit can manipulate and transiently represent information from spatial
WM; however, we do not claim that the original representation of object
location, maintained in DLPFC, is necessarily altered by this procedure.
Therefore, the original WM snapshot information is still available for future
transformation and retrieval. This implies that subjects are still more accu-
rate when tested at a presentation viewpoint because no transformation is
required and no error is introduced. Our model does not, however, address
the question of why information from a particular presentation viewpoint
is maintained in preference to others.

Our model also provides an explanation for the Wang and Spelke (2000)
data. If we assume that subjects have an accurate WM snapshot of object lo-
cations before disorientation, then after disorientation, they could mentally
navigate from the original snapshot viewpoint to their new estimated view-
point while simultaneously using the egomotion signal that drives mental
navigation to transform egocentric object location coordinates. From the
discussion, we believe that the mental navigation and parietal update pro-
cedure would introduce error into an egocentric object location transforma-
tion. Since we have assumed serial update, that is, the mental navigation
procedure must be performed separately for each object, the independent
error inflicted on each object coordinate transformation could generate an
overall object configuration error. This would also explain the object con-
figuration error obtained from the experiment in which subjects were reori-
ented after disorientation by a bright light visible through their translucent
blindfold. In this case, a mental navigation still must be performed from the
viewpoint of the subjects while facing the light directly to the new view-
point.

Wang and Spelke’s (2000) results indicate that blindfolded subjects who
are continuously oriented are able to maintain object configurations accu-
rately even after rotation. Performance in these conditions, which include
slow rotation or rotation in the presence of an orienting light, suggests min-
imal transformation of internal object location representations, or transforma-
tion that induces a systematic error only. One possible explanation for
how subjects could recall object locations under continuously oriented con-
ditions is that they could generate a displacement-rotation vector via path integration. This vector could simply be added to the DLPCF-maintained object location representations at each recall. In this way, all object locations would be subject to the same systematic error. Another possibility is that the subject who remains oriented makes use of some allocentric system in which object locations are static. In this case, the subject must simply orient himself or herself with regard to the configuration. Given the availability of path integration in these conditions, subjects would be able to accurately localize themselves within the environment, and this procedure should result only in a small error, which would be the same for all objects.

In addition to being consistent with a range of experimental findings regarding memory for object locations, our model makes two easily testable predictions. First, the serial updating we require for multiple objects suggests that a subject’s reaction time to determine if an object configuration, seen from a novel viewpoint, is the same as one seen from a different viewpoint should monotonically increase with number of objects. Second, our hypothesis regarding how object configuration error is introduced into the Wang and Spelke (2000) task suggests that as subjects become more familiar with a given environment, and hence develop a more precise cognitive map, their performance on these types of tasks should improve. These predictions are currently being investigated empirically. Should our first prediction turn out to be false, it may be that our model is too simple. A more complicated model, in which object locations are updated in parallel and egocentric space is represented nonuniformly with respect to the main layer neuron density, might also be able to account for the object configuration error observed by Wang and Spelke.

Finally, another interesting question to investigate would be the nature of how objects become represented after being seen as stable for long periods of time, that is, after several minutes. For example, the work of Jeffery (1998) suggests that objects that remain stable over time begin to affect place cell firing in rats. Hence, objects might become part of the animal’s cognitive map after some time. One way to investigate this possibility would be to perform an experiment similar to that of Wang and Spelke (2000) in which the presentation phase was continued over a long period of time before testing began. It should be noted, however, that Cressant, Muller, and Poucet (1997) found that objects near the center of an environment failed to affect place cell firing in the same way. Therefore, such results would likely be found only for certain object placements.

Acknowledgments

This research was supported by a grant from the Natural Sciences and Engineering Research Council of Canada to S.B. We thank Neil Burgess for helpful discussions.
References


Received September 16, 2003; accepted February 6, 2004.
This article has been cited by:

