Remapping attention in multiple object tracking

Piers D.L. Howe [a,*], Trafton Drew [b,c], Yair Pinto [d], Todd S. Horowitz [b,c]

[a] University of Melbourne, Melbourne, Australia
[b] Brigham and Women's Hospital, Boston, USA
[c] Harvard Medical School, Boston, USA
[d] University of Amsterdam, Amsterdam, The Netherlands

ARTICLE INFO

ABSTRACT

Which coordinate system do we use to track moving objects? In a previous study using smooth pursuit eye movements, we argued that targets are tracked in both retinal (retinotopic) and scene-centered (allocentric) coordinates (Howe, Pinto, & Horowitz, 2010). However, multiple object tracking typically also elicits saccadic eye movements, which may change how object locations are represented. Observers fixated a cross while tracking three targets out of six identical disks confined to move within an imaginary square. The fixation cross alternated between two locations, requiring observers to make repeated saccades. By moving (or not moving) the imaginary square in sync with the fixation cross, we could disrupt either (or both) coordinate systems. Surprisingly, tracking performance was much worse when the objects moved with the fixation cross, although this manipulation preserved the retinal image across saccades, thereby avoiding the visual disruptions normally associated with saccades. Instead, tracking performance was best when the allocentric coordinate system was preserved, suggesting that targets locations are maintained in that coordinate system across saccades. This is consistent with a theoretical framework in which the positions of a small set of attentional pointers are predictively updated in advance of a saccade.

© 2011 Elsevier Ltd. All rights reserved.

1. Introduction

Visual information enters the brain via the retina, so our initial representation of the world is intrinsically retinotopic. Our retinas, however, are moving up to three times every second (Burr, 2004) ensuring that the retinal image is constantly changing. Yet, we manage to maintain both the perceptual impression of a stable world and the ability to interact with one. The problem of how this feat is accomplished has been the subject of a great deal of work in the field of perception (e.g. Burr, 2004; Cavanagh, Hunt, Afraz, & Rolfs, 2010; d'Avossa et al., 2007; Duhamel, Colby, & Goldberg, 1992; Irwin, 1996; Melcher, 2007, 2008; Wurtz, 2008). This problem is compounded once objects in the world are also moving independently of the eyes (and/or the observer is moving with respect to the world).

Consider the case of multiple object tracking (MOT), a task originally devised by Pylyshyn and Storm (1988). MOT is basically a laboratory-based "shell game": observers are presented with an array of identical objects, a subset of which are designated as the "targets". The observer's task is to track these targets as they move

independently and unpredictably for some time, and then point out which items correspond to the original targets. If there were a single target, an observer might simply track it with pursuit eye movements, but since there are several targets moving in different directions, this strategy will fail, and the assumption is that observers must track the targets with either multi-focal attention (Cavanagh & Alvarez, 2005) or at least pre-attentive indexes (Pylyshyn, 2007).

While the task cannot be solved with simple pursuit eye movements, observers tend to move their eyes while tracking (Fehd & Seiffert, 2008, 2010; Huff, Papenmeier, Jahn, & Hess, 2010b; Landry, Sheridan, & Yufik, 2001; Zelinsky & Neider, 2008). Indeed, requiring observers to fixate reduces performance (Intriligator & Cavanagh, 2001). However, it may be that it is the effort required to fixate, rather than the absence of eye movements, that impairs tracking. One might wonder why observers would want to make saccades at all during tracking given that they disrupt the retinotopic frame of reference. One possibility is that the advantage of using the high resolution fovea to resolve potential collisions (Zelinsky & Todor, 2010) outweighs the temporary disruption caused by the saccade.

How can observers continue to track objects despite these eye movements? One way to account for this is to assume that MOT is carried out in an allocentric representation independent of the retina. For example, Liu and her colleagues (2005) asked observers

* Corresponding author.
E-mail addresses: PiersDouglasHowe@gmail.com (P.D.L. Howe), tdrew1@rics.bwh.harvard.edu (T. Drew), yair.pinto@gmail.com (Y. Pinto), toddh@search.bwh.harvard.edu (T.S. Horowitz).

to track targets within a simulated 3D volume on a computer monitor. The volume containing the tracking stimuli then moved unpredictably, not only translating across the screen but also rotating and appearing to move closer to or further from the viewer. Their observers had little trouble with this task. In fact, it was not until they projected the MOT stimuli onto a convex surface that they were able to disrupt performance. They suggested that the visual system represented the tracking stimuli in an "object-centered" representation, relative to the containing volume (a 3D wireframe). The account is supported by work from Huff, Meyerhoff, Papenmeier, and Jahn (2010a) who demonstrated that observers could successfully track items that disappeared during viewpoint changes, as long as the changes themselves were continuous rather than abrupt.

Interestingly, neither of these studies took eye movements into account. In a recent study, we directly contrasted retinotopic and allocentric reference frames by manipulating observers eye movements during MOT (Howe et al., 2010). We asked observers to fixate on a cross while tracking targets contained within an imaginary square centered a fixed distance off the fixation cross. We constructed five conditions in which we moved either the imaginary square, the fixation cross, both or neither, in such a way that we independently varied whether the retinotopic or allocentric coordinates of the tracked objects were preserved, while also controlling for the difficulty of pursuing a moving fixation cross. We found that moving the imaginary square in *either* coordinate system was sufficient to impair tracking performance, but disrupting *both* coordinate systems simultaneously had only little additional effect on tracking performance. Consequently, we proposed that objects are tracked in both retinotopic and allocentric coordinate systems, and that proper coordination between the two is required for successful tracking.

Howe et al. (2010) was exclusively concerned with smooth pursuit eye movements, in which the relationship between reference frames changed smoothly and continuously, and the stimulus was continuously visible. As we previously noted, under free viewing conditions observers typically use a mixture of smooth pursuit and saccadic eye movements, sometimes fixating the centroid of the targets, sometimes fixating or pursing individual items, using a mix of strategies that depends on the individual and on the tracking load (Fehd & Seiffert, 2008, 2010; Huff et al., 2010b; Landry et al., 2001; Zelinsky & Neider, 2008). Huff et al. (2010a) showed that abrupt translations of a tracking display inhibit tracking more than smooth translations. Naturalistic tracking typically involves abrupt shifts of eye position, which would cause abrupt changes of the retinal image. Compensating for these disruptions therefore may require a different computational strategy to that employed when observers track objects using smooth eye pursuits.

Seiffert (2005) studied how saccadic eye movements affect MOT performance. In her study, the tracking stimuli either moved as a group relative to the fixation point, or the fixation point was moved relative to the tracking stimuli. In the first situation, the speeds of the stimuli were increased in both the retinotopic and allocentric coordinate systems, conversely in the second situation the speeds of the stimuli increased only in the retinotopic representation. Because these two manipulations had a similar detrimental effect on tracking accuracy, Seiffert concluded that tracking must occur in the retinotopic coordinate system.

Note that the Seiffert (2005) study compared a condition where both coordinate systems were disrupted to one where only the retinotopic coordinate system was disrupted. Our approach was to also consider a third condition where the allocentric coordinate system was disrupted but the retinotopic coordinate system preserved. By considering this new condition, as well as the two conditions considered by Seiffert (2005), we could address the possibility that disruption of either coordinate system is sufficient to disrupt tracking, as observed in our previous study (Howe et al., 2010). In addition, we were able to ensure that the mental load of fixating on the fixation cross was the same in all three conditions. Observers performed an MOT task on stimuli that moved within an imaginary square. We directly manipulated the point of fixation and the position of the imaginary square to create three conditions. In all conditions, observers made exactly the same eye movements, with the fixation cross shifting back and forth between two positions horizontally separated by 4.5°. In the *allocentric-preserved* condition, the imaginary square remained stationary on the screen. The stimuli were thus stable with respect to an allocentric representation, while shifting abruptly back and forth on the retina, thereby disrupting the retinotopic coordinate system (see left column of Fig. 1). In the *retinotopic-preserved* condition, we moved the imaginary square in sync with the eyes, so that the stimuli were stable on the retina, but shifting abruptly in the world, thereby disrupting the allocentric coordinate system (see middle column of Fig. 1). Finally, in the *neither-preserved* condition, we moved the imaginary square such that it shifted in both reference frames (see right column of Fig. 1).

Based on the literature, we anticipated three possible outcomes from this design. First, our prior work with smooth pursuit eye movements suggests that target locations are tracked in both reference frames (Howe et al., 2010). From this perspective, we would predict that tracking performance will be roughly equivalent in all three conditions.

However, as Huff et al. (2010a) pointed out, abrupt changes to the coordinate systems may be handled differently from smooth changes. When they induced abrupt viewpoint changes, they found data more supportive of a retinotopic representation for tracking (see also Huff, Jahn, & Schwan, 2009; Seiffert, 2005). Furthermore, a compelling set of studies by Golomb and her colleagues suggests the locus of attention is actually updated more slowly than the locations of visual stimuli following a saccade, leaving behind a trace in retinotopic coordinates which can be detected in both behavioral (Golomb, Chun, & Mazer, 2008) and neural measures (Golomb, Nguyen-Phuc, Mazer, McCarthy, & Chun, 2010). On this account, in the *allocentric-preserved* condition, it might take a couple hundred ms for attention to catch up to the target locations after a saccade, while in the *retinotopic-preserved* case attention would always be in the right place. Both Huff and Golomb's accounts seem to predict an advantage in the *retinotopic-preserved* condition.

This prediction is especially plausible as the *retinotopic-preserved* condition avoids the major problem associate with saccades: the fact that saccades change the retinal image. In the *retinotopic-preserved* condition the tracked objects move in sync with the observer's fixation, thereby preserving the retinal image across saccades and avoiding the visual disruptions that typically accompany saccades. This obviates the need for saccadic remapping of the retinotopic locations of the tracked objects. This alone would suggest that tracking should be easiest in the *retinotopic-preserved* condition.

Finally, one could predict that the location of tracked targets will be represented allocentrically across saccades. Wurtz (2008) has proposed a sparse theory of saccadic remapping, in which only a small set of relevant, attended stimuli have their positions updated across saccades. This theory has the advantage of accounting for the apparent stability of the world while limiting the computational complexity involved in remapping the entire visual field. Cavanagh et al. (2010) suggest that a small set of attentional pointers are predictively updated in advance of a saccade. Their system of attentional pointers is quite similar to Pylyshyn's FINST (Fingers of INSTantiation, 1989) account of MOT, and indeed Pylyshyn has suggested that FINSTs also provide a sparse mechanism for translating the coordinates of relevant objects between different
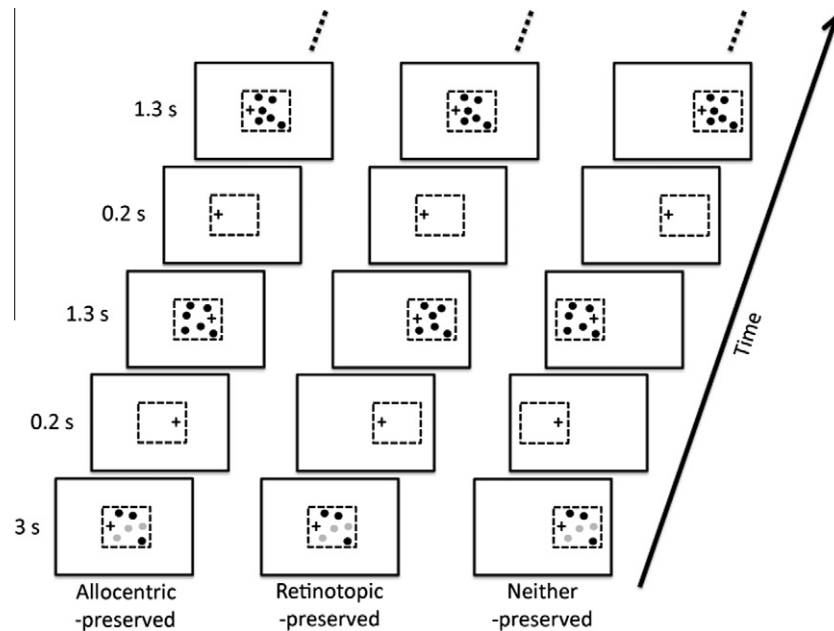
Fig. 1. Schematics of the stimuli used in Experiment 1. There were three conditions, named after which coordinate system was preserved. Whenever, the observer made a saccade, the stimuli disappeared for 200 ms. Please see Section 2.1 for further details.

reference frames (Pylyshyn, 2007). In this view, then, tracking is accomplished by assigning attentional pointers to targets, and the positions of these pointers are predictively remapped just before a saccade is executed, yielding a prediction of superior performance in the *allocentric-preserved* condition.

To preview our results, across three experiments, we found a substantial advantage for the *allocentric-preserved* condition, in line with the predictions from the sparse predictive remapping framework (Cavanagh et al., 2010; Pylyshyn, 2007; Wurtz, 2008). We conclude by discussing the implications of this finding.

2. Experiments

2.1. Experiment 1

In this experiment, observers tracked three of six disks that were confined to move within an imaginary square. The fixation cross alternated between one of two positions, forcing the observer to make repeated saccades. In the *retinotopic-preserved* condition, the imaginary square (and thus the disks contained within it) would move with the fixation cross so as to preserved the retinotopic coordinates of the disks, while disrupting their allocentric coordinates. In the *allocentric-preserved* condition, the imaginary square remained stationary so as to preserved the allocentric coordinates of the disks, while disrupting their retinotopic coordinates. In the *neither-preserved* condition, the imaginary square moved so as to disrupt both the allocentric and retinotopic coordinates of the disks.

2.1.1. Observers

Twelve paid volunteers participated in the experiment. All participants had normal or corrected-to-normal vision. The mean age was 25.1 years.

2.1.2. Apparatus and stimuli

The stimuli were presented on a 21-in Mitsubishi Diamond Pro CRT monitor at a refresh rate of 75 Hz using the Psychophysics Toolbox (version 3) for MATLAB® (Brainard, 1997; Pelli, 1997). The stimuli are shown in Fig. 1. The tracking stimuli comprised six black disks on a white background. The disks were confined to move within the confines of an imaginary square, whose boundaries were invisible. Each disk subtended 1° of visual angle and the imaginary square subtended 15°. The disks moved in straight lines except when they bounced off each other or the sides of the imaginary square. The disks were surrounded by imaginary buffers that ensured that the center-to-center distance of any two disks could never be less than 3°.

Fixation was enforced using an Arrington Research eye tracker. Immediately after the fixation cross changed location, the observers were given a 0.5 s grace period to establish fixation on the new location of the fixation cross. If after this time fixation deviated by more than 2° from the fixation cross, the trial was aborted and redone. The eye tracker was recalibrated after every 40 trials, or sooner if there was any evidence that it had become uncalibrated, such as repeated fixation errors.

2.1.3. Procedure

In all conditions, the fixation cross could occupy one of two locations horizontally offset by 4.5° either to the left or to the right of the center of the computer monitor. During the trial, the fixation cross would alternate between these two positions every 1.5 s. When the observer initiated the saccade from the old fixation location to the new fixation location the disks would disappear for 200 ms (the blank interval). This time interval ensured that the observer had finished the saccade before the disks reappeared, as verified in a pilot study using the eye tracker. During this delay period, the motion of the dots ceased, i.e. they did not move while they were invisible.

In the *allocentric-preserved* condition, the imaginary square remained stationary. This preserved the disks' allocentric coordinates, but not their retinotopic coordinates. Converse, in the *retinotopic-preserved* condition the imaginary square moved with the fixation cross, so as to preserve the retinotopic coordinates of the disks, at the expense of their allocentric coordinates. Thus, the fixation cross occupied the same two positions as it did in the *allocentric-preserved* condition, but the imaginary square occupied different positions, because it was forced to move with the fixation cross.

In the *neither-preserved* condition, the fixation cross again occupied the same two positions as it did in the other two conditions. Now, however, the imaginary square would move so that neither the allocentric or the retinotopic locations of the disks were preserved across saccades (c.f. Fig. 1). When the fixation cross occupied its leftmost position, the imaginary square was horizontally offset 4.5° to the left of the fixation cross. Conversely, when the fixation cross occupied its rightmost position, the imaginary square was then offset 4.5° to the right of the fixation cross. In this way the average distance of any one object from the fixation cross was the same in all three conditions.

At the start of the trial, three of the disks would turn red for 3 s to indicate that they were the targets to be tracked. Then, they would revert to black and continue to move within the confines of the imaginary square for a further 5 s. At the end of the trial, the observer would use the mouse to indicate the target disks. The trial was counted as correct only if the observer correctly identified all three target disks.

For each condition, thirty trials were run and the QUEST routine (King-Smith, Grigsby, Vingrys, Benes, & Supowit, 1994; Watson & Pelli, 1983) was used to find the disk speed that resulted in all three targets being identified correctly on 70% of the trials. The order of conditions was counterbalanced across observers.

2.1.4. Results

We normalized the speed thresholds for each observer to the speed obtained for that observer in the *allocentric-preserved* condition. This ensured that the data from each observer had approximately the same variance, allowing us to average across observers. The mean threshold in the *allocentric-preserved* condition was 13.5°/s. Fig. 2 shows the normalized data averaged across observers for all three experiments.

Tracking was easiest when the stimuli remained fixed in the world, even if they moved on the retina. We found a substantial advantage for the *allocentric-preserved* condition over the *retinotopic-preserved* ($t(11) = 7.27$, $p < 0.001$) and *neither-preserved* conditions ($t(11) = 21.7$, $p < 0.001$). This suggests that observer observers use a allocentric coordinate system to track objects, at least in the presence of saccades.

2.2. Experiment 2

In Experiment 1 there was a 200 ms blank interval after the initiation of each saccade. During this interval the disks were not visible. This was to ensure that each saccade had ended before the

disks reappeared. However, an unintended consequence of the blank interval is that it may have allowed sufficient time for the retinotopic attentional trace to fade (Golomb et al., 2008). It could be that, immediately after the saccade, the disks were represented retinotopically, but because we waited 200 ms before displaying the disks, the observer could not take advantage of this retinotopic representation. If so, then including a blank interval may have inadvertently favored an allocentric representation. Experiment 2 was designed to test this hypothesis. Specifically, we investigated whether the advantage would shift towards the *retinotopic-preserved* condition if there was no blank interval.

2.2.1. Stimuli and procedure

Other than causing the disks not to disappear during the saccades, the stimuli and procedure for Experiment 2 were identical to that of Experiment 1 (c.f. Fig. 3). As before 12, paid observers were used, six of which had participated in Experiment 1. Their mean age was 29.8 years.

2.2.2. Results

As with Experiment 1, we normalized the speed thresholds for each observer to the speed obtained for that observer in the *allocentric-preserved* condition. The mean threshold in the *allocentric-preserved* condition was again 13.5°/s. Fig. 2 shows the normalized data averaged across the observers.

Performance was still much better in the *allocentric-preserved* condition than in the *retinotopic-preserved* condition ($t(11) = 52.6$, $p < 0.001$). In fact, performance in the *retinotopic-preserved* condition in Experiment 2 was significantly less than that in Experiment 1 ($t(22) = 4.21$, $p < 0.001$).

2.3. Experiment 3

In both Experiments 1 and 2, performance in the *neither-preserved* condition was significantly less than that in *retinotopic-preserved* condition (1, $t(11) = 2.73$, $p = 0.02$; 2, $t(11) = 2.97$, $p = 0.013$). This raises the possibility that tracking uses a retinotopic representation in addition to an allocentric representation (c.f. Howe et al., 2010). However, in the *neither-preserved* condition, the imaginary square moved over a greater (screen) distance than it did in the *retinotopic-preserved* condition, which may have hampered performance in the *neither-preserved* condition, especially if the dominant coordinate system is allocentric, as our data suggests. Experiment 3 addressed this potential confound by rearranging the stimuli so that the imaginary square traveled the same distance in both conditions (Fig. 4).

2.3.1. Stimuli and procedure

The procedure was identical to Experiment 2 and, again, 12 paid observers were used. Their mean age was 27.0 years. Six had participated in Experiment 2 and seven had participated in Experiment 1.

As before, the fixation cross could occupy one of two locations horizontally offset by 4.5° either to the left or to the right of the center of the computer monitor. However, whereas in Experiment 2 the imaginary square was vertically centered on the fixation cross, in Experiment 3 the center of the imaginary square was offset by 4.5° horizontally and 4.5° vertically from the fixation cross.

As before, in the *allocentric-preserved* condition the imaginary square remained stationary throughout the trial, so as to preserve the allocentric coordinates of the disks. Conversely, in the *retinotopic-preserved* condition the imaginary square moved with the observer's fixation so as to preserve the retinotopic coordinates of the disks. In the *neither-preserved* condition, the imaginary square would alternate from being above and below the fixation cross. This ensured that its total displacement (over the computer
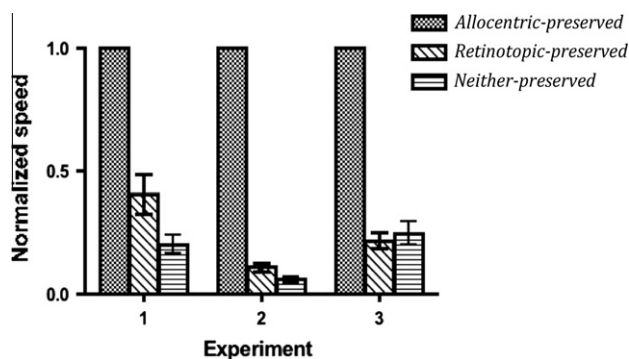

Fig. 2. Normalized speed thresholds. In each condition, we obtained the tracking speed that resulted in the observer identifying all targets correctly on 70% of the trials. For each observer, these thresholds were then divided by that observer's performance in the *allocentric-preserved* condition to obtain the normalized speed. Error bars represent standard errors of the mean. Please see Section 2 for further details.
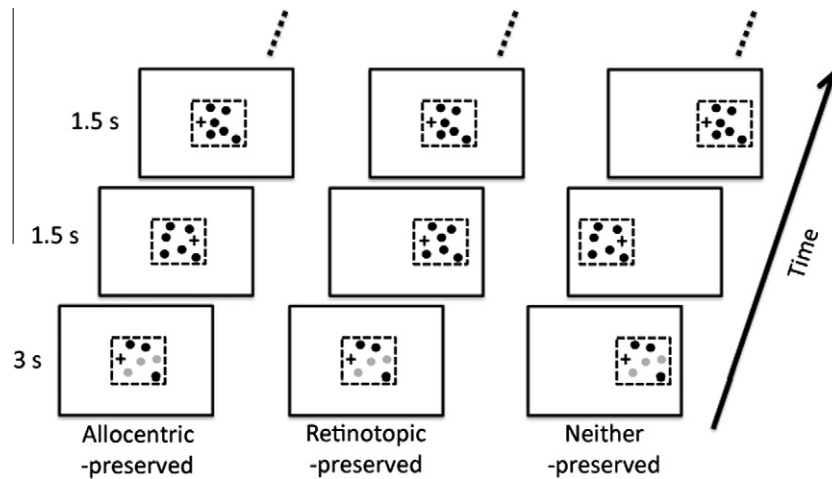
Fig. 3. Schematics of the stimuli used in Experiment 2. This experiment was identical to the first experiment except that the stimuli did not disappear whenever the observer made a saccade. Please see Section 2.2 for further details.
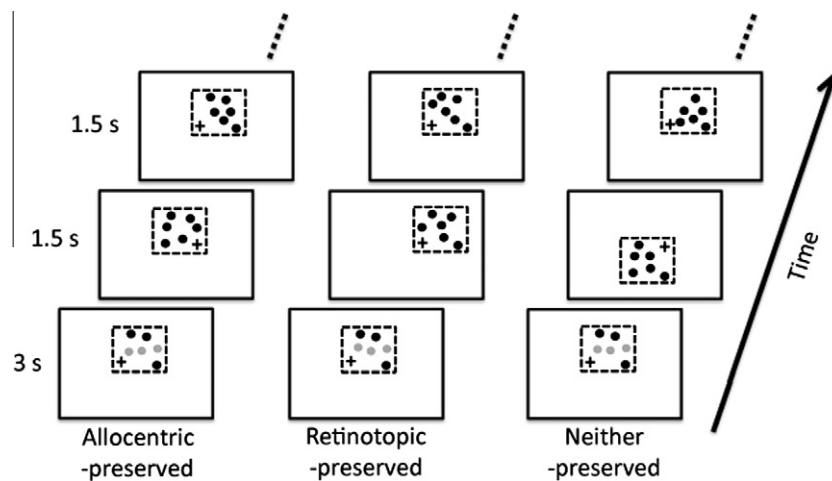

Fig. 4. Schematics of the stimuli used in Experiment 3. This experiment was identical to Experiment 2, except that during each saccade the disks were displaced the same distance in the *neither-preserved* condition as in the *retinotopic-preserved* condition. Please see Section 2.3 for further details.

monitor) was the same in the *neither-preserved* condition as it was in the *retinotopic-preserved* condition (9°). This allowed us to fairly compare the performance in the two conditions.

2.3.2. Results

As with Experiment 2, we normalized the speed thresholds for each observer to the speed obtained for that observer in the *allocentric-preserved* condition. The mean threshold in the *allocentric-preserved* condition was 9.9°/s. This value is considerably lower than that observed in the other two experiment (in both cases 13.5°/s). This is because the task was slightly harder in Experiment 3, consistent with the fact that in this experiment the disks were on average slightly further from the point of fixation, which would be expected to make tracking more difficult (Intriligator & Cavanagh, 2001).

We again replicated the advantage of the *allocentric-preserved* condition over the *retinotopic-preserved* condition ($t(11) = 24.5$, $p < 0.001$). However, with the displacement of the imaginary square controlled (i.e. the same in both the *retinotopic-preserved* condition and the *neither-preserved* condition), there was no evidence that tracking was worse in the *neither-preserved* condition than in the *retinotopic-preserved*. In fact, performance was slightly greater in this condition than in the *retinotopic-preserved* condition,

though this was not statistically significant ($t(11) = 0.858$, $p = 0.409$).

3. Discussion

In all three experiments, performance was greatest in the *allocentric-preserved* condition. Performance in the *retinotopic-preserved* condition was much worse and proved to be no better than performance in the *neither-preserved* condition, once the displacement of the imaginary square was equated across both conditions. This result is consistent with the sparse remapping hypothesis. Perceptual stability is maintained across saccades by selectively remapping attentional pointers (Cavanagh et al., 2010; Pylyshyn, 2007; Wurtz, 2008). Since these pointers are also used to track targets in MOT, the locations of targets in MOT are automatically remapped; if remapping were optional or strategic, observers would have adapted to the *retinotopic-preserved* condition by inhibiting the remapping. Having the targets reappear at their retinotopic locations therefore increased tracking difficulty, although this difficulty can be somewhat mitigated if the targets do not reappear immediately after the termination of the saccade, as evidenced by tracking performance being greater in the *retinotopic-preserved* condition of Experiment 1 than in the same

condition of Experiment 2. Presumably, delaying the reappearance of the targets in Experiment 1 gave the visual system more time to recalibrate where it expected the targets to reappear.

How does this relate to our previous finding that disrupting either coordinate system impairs tracking (Howe et al., 2010)? The reason for the discrepancy between the two studies is unclear, but one possibility is as follows. Howe et al. (2010) used a very similar set of stimuli to the present study except that both the fixation cross and the imaginary square moved in smooth fashion. The stimuli were continuously visible and there were no abrupt shifts in either coordinate system. In contrast, in the present study observers made saccades. Saccadic suppression substantially reduces the visibility of stimuli (Matin, 1974), particularly interfering with location information (Irwin & Brockmole, 2004). Under these conditions, the relationship between the two coordinate systems appears to be altered. Retinotopic activity is shifted prospectively to match the expected locations of the targets after the saccade (Duhamel et al., 1992). We speculate that this allowed the tracking system to recover from disruptions of the retinotopic coordinate system, explaining why only disruptions of the allocentric coordinate system were able to disrupt tracking in the present study.

While the above paragraph was speculative in nature, there is ample evidence supporting its major assumption that neural activity can remapped across saccades (Duhamel et al., 1992). This has been demonstrated in a wide variety of tasks (Melcher, 2005; Melcher, 2007, 2008). Perhaps most importantly from the perspective of MOT, remapping appears to occur also in the motion domain (Melcher & Morrone, 2003). Consistent with Melcher and Morrone (2003), an fMRI investigation has demonstrated spatiotopic mapping of the human middle temporal cortex (MT; d'Avossa et al., 2007), an area that is clearly implicated in motion processing in general and in multiple object tracking in particular (Culham, Cavanagh, & Kanwisher, 2001; Culham et al., 1998; Howe, Horowitz, Morocz, Wolfe, & Livingstone, 2009; Jovicich et al., 2001). Interestingly, this spatiotopic mapping reverted to retinotopic mapping when the observer's attention was diverted (Gardner, Merriam, Movshon, & Heeger, 2008), suggesting that attention is needed to maintain a spatiotopic representation (Burr et al., 2010; Crespi et al., 2009), consistent with the sparse remapping hypothesis.

While in our experiments we always found tracking accuracy to be greatest when the allocentric coordinate system was preserved, this need not have been the case. We conclude by revisiting the three rationales predicting a retinotopic preference, and suggest how the results we report here may be reconciled with previous studies. The studies by Huff and colleagues (Huff et al., 2009, 2010a) produced results consistent with retinotopy when there were abrupt transitions in their displays, but not when there were smooth transitions. As our experiments involved saccades, which in turn engendered abrupt transitions, one might have expected our data to favor a retinotopic representation. A possible reason why we obtained a different result to that reported by Huff et al. is that the proximal cause for the abrupt transitions in our experiments was the visual system itself, which has a built-in mechanism for compensating for such transitions (Duhamel et al., 1992). Compensating for an externally-generated rotation of an external coordinate system, such as that found in the Huff et al. experiments, is presumably much more problematic. We suggest that this is why Huff et al. observed that abrupt disruptions of the retinotopic coordinate system disrupted tracking, leading them to concluding that tracking must therefore occur in retinotopic coordinates.

Seiffert (2005) also considered a condition where the retinotopic coordinate system was disrupted by saccades, while preserving the allocentric coordinate system. In contrast to our findings, Seiffert found that this manipulation disrupted tracking as much as the condition where both coordinate systems were disrupted, leading her to conclude that tracking must therefore occur in retinotopic coordinates. We suggest that the key difference between her experiment and ours is that in all of our conditions the fixation load was held constant. Specifically, in all our experiments, the observer always had to track a fixation cross that alternated between two locations. Conversely, in the Seiffert (2005) *both-disrupted* condition, the fixation cross was stationary whereas in her *allocentric-preserved* condition the fixation cross was moving. Fixating a moving cross is more difficult than fixating a stationary one (Howe et al., 2010). This could explain why performance in her *allocentric-preserved* condition was not greater than that in her *both-disrupted* condition.

Our findings also appear to conflict with Golomb and colleagues recent suggestion that the native coordinate system of attention is retinotopic (Golomb et al., 2008, 2010). Note that in the critical conditions of the Golomb paradigm (see also Mathôt & Theeuwes, 2010) observers were instructed to attend to a blank location, make a saccade, and then to report the property of a probe object which appears at either the allocentric or retinotopic correspondent of the original location. Thus, attention is directed to a *location*, but not to an existing *object*. In contrast, in our paradigm, attention is directed to objects that are visible before and after the saccade. Following Cavanagh et al. (2010), we suggest that when attention is directed to an object, the retinotopic locus of attention is shifted automatically when the object is remapped. When attention is directed to an empty location in space, however, there is no object being remapped, so shifting the retinotopic locus may take longer. Along a similar line, during search attention can shift more quickly with a moving object than between objects (Horowitz, Holcombe, Wolfe, Arsenio, & DiMase, 2004; Verstraten, Cavanagh, & Labianca, 2000). Thus, the object of attention serves as anchor, not just for the perception of a stable visual world, but also for attention itself.

Acknowledgments

References

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10*(4), 433–436.
Burr, D. (2004). Eye movements: Keeping vision stable. *Current Biology, 14*(5), R195–197.
Burr, D., Crespi, S., Biagi, L., Tosetti, M., d'Avossa, G., & Morrone, M. C. (2010). *Spatiotopic coding of BOLD signal in visual cortex depends on spatial attention. Neuroscience meeting planner.* San Diego, CA: Society for Neuroscience.
Cavanagh, P., & Alvarez, G. A. (2005). Tracking multiple targets with multifocal attention. *Trends in Cognitive Sciences, 9*(7), 349–354.
Cavanagh, P., Hunt, A. R., Afraz, A., & Rolfs, M. (2010). Visual stability based on remapping of attention pointers. *Trends in Cognitive Sciences, 14*(4), 147–153.
Crespi, S., Biagi, L., Burr, D., d'Avossa, G., Tosetti, M., & Morrone, M. C. (2009). Spatial attention modulates the spatiotopicity of human MT complex. *Perception, 38*(ECVP Abstract Suppl.), 6.
Culham, J. C., Brandt, S. A., Cavanagh, P., Kanwisher, N. G., Dale, A. M., & Tootell, R. B. (1998). Cortical fMRI activation produced by attentive tracking of moving targets. *Journal of Neurophysiology, 80*(5), 2657–2670.
Culham, J. C., Cavanagh, P., & Kanwisher, N. G. (2001). Attention response functions: Characterizing brain areas using fMRI activation during parametric variations of attentional load. *Neuron, 32*(4), 737–745.
d'Avossa, G., Tosetti, M., Crespi, S., Biagi, L., Burr, D. C., & Morrone, M. C. (2007). Spatiotopic selectivity of BOLD responses to visual motion in human area MT. *Nature Neuroscience, 10*(2), 249–255.
Duhamel, J. R., Colby, C. L., & Goldberg, M. E. (1992). The updating of the representation of visual space in parietal cortex by intended eye movements. *Science, 255*(5040), 90–92.
Fehd, H. M., & Seiffert, A. E. (2008). Eye movements during multiple object tracking: Where do participants look? *Cognition, 108*(1), 201–209.
Fehd, H. M., & Seiffert, A. E. (2010). Looking at the center of the targets helps multiple object tracking. *Journal of vision, 10*(4), 11–13 (19).

Gardner, J. L., Merriam, E. P., Movshon, J. A., & Heeger, D. J. (2008). Maps of visual space in human occipital cortex are retinotopic, not spatiotopic. *Journal of Neuroscience, 28*(15), 3988–3999.

Golomb, J. D., Chun, M. M., & Mazer, J. A. (2008). The native coordinate system of spatial attention is retinotopic. *Journal of Neuroscience, 28*(42), 10654–10662.

Golomb, J. D., Nguyen-Phuc, A. Y., Mazer, J. A., McCarthy, G., & Chun, M. M. (2010). Attentional facilitation throughout human visual cortex lingers in retinotopic coordinates after eye movements. *Journal of Neuroscience, 30*(31), 10493–10506.

Horowitz, T. S., Holcombe, A. O., Wolfe, J. M., Arsenio, H. C., & DiMase, J. S. (2004). Attentional pursuit is faster than attentional saccade. *Journal of vision, 4*(7), 585–603.

Howe, P. D., Horowitz, T. S., Morocz, I. A., Wolfe, J. M., & Livingstone, M. S. (2009). Using fMRI to distinguish components of the multiple object tracking task. *Journal of Vision, 9*, 1–11.

Howe, P. D., Pinto, Y., & Horowitz, T. S. (2010). The coordinate systems used in visual tracking. *Vision Research, 50*, 2375–2380.

Huff, M., Jahn, G., & Schwan, S. (2009). Tracking multiple objects across abrupt viewpoint changes. *Visual Cognition, 17*, 297–306.

Huff, M., Meyerhoff, H. S., Papenmeier, F., & Jahn, G. (2010a). Spatial updating of dynamic scenes: Tracking multiple invisible objects across viewpoint changes. *Attention, Perception and Psychophysics, 72*(3), 628–636.

Huff, M., Papenmeier, F., Jahn, G., & Hess, R. F. (2010b). Eye movements across viewpoint changes in multiple object tracking. *Visual Cognition, 18*(9), 1368–1391.

Intriligator, J., & Cavanagh, P. (2001). The spatial resolution of visual attention. *Cognitive Psychology, 43*(3), 171–216.

Irwin, D. E. (1996). Integrating information across saccadic eye movements. *Current Directions in Psychological Science, 5*, 94–100.

Irwin, D. E., & Brockmole, J. R. (2004). Suppressing where but not what: The effect of saccades on dorsal- and ventral-stream visual processing. *Psychological Science, 15*(7), 467–473.

Jovicich, J., Peters, R. J., Koch, C., Braun, J., Chang, L., & Ernst, T. (2001). Brain areas specific for attentional load in a motion-tracking task. *Journal of Cognitive Neuroscience, 13*(8), 1048–1058.

King-Smith, P. E., Grigsby, S. S., Vingrys, A. J., Benes, S. C., & Supowit, A. (1994). Efficient and unbiased modifications of the QUEST threshold method: Theory, simulations, experimental evaluation and practical implementation. *Vision Research, 34*(7), 885–912.

Landry, S., Sheridan, T., & Yufik, Y. M. (2001). A methodology for studying cognitive groupings in a target-tracking task. *IEEE Transactions on Intelligent Transportation Systems, 2*(2), 92–100.

Liu, G., Austen, E. L., Booth, K. S., Fisher, B. D., Argue, R., Rempel, M. I., et al. (2005). Multiple-object tracking is based on scene, not retinal, coordinates. *Journal of Experimental Psychology: Human Perception and Performance, 31*(2), 235–247.

Mathôt, S., & Theeuwes, J. (2010). Evidence for the predictive remapping of visual attention. *Experimental Brain Research, 200*(1), 117–122.

Matin, E. (1974). Saccadic suppression: A review and an analysis. *Psychological Bulletin, 81*(12), 899–917.

Melcher, D. (2005). Spatiotopic transfer of visual-form adaptation across saccadic eye movements. *Current Biology, 15*(19), 1745–1748.

Melcher, D. (2007). Predictive remapping of visual features precedes saccadic eye movements. *Nature Neuroscience, 10*(7), 903–907.

Melcher, D. (2008). Dynamic, object-based remapping of visual features in trans-saccadic perception. *Journal of Vision, 8*(14), 1–17.

Melcher, D., & Morrone, M. C. (2003). Spatiotopic temporal integration of visual motion across saccadic eye movements. *Nature Neuroscience, 6*(8), 877–881.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10*(4), 437–442.

Pylyshyn, Z. W. (1989). The role of location indexes in spatial perception: A sketch of the FINST spatial-index model. *Cognition, 32*(1), 65–97.

Pylyshyn, Z. W. (2007). *Things and places: How the mind connects with the world*. MA, MIT Press: Cambridge.

Pylyshyn, Z. W., & Storm, R. W. (1988). Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial Vision, 3*(3), 179–197.

Seiffert, A. E. (2005). Attentional tracking across display translations. *Journal of Vision, 5*(8), 643a (abstract).

Verstraten, F. A., Cavanagh, P., & Labianca, A. T. (2000). Limits of attentive tracking reveal temporal properties of attention. *Vision Research, 40*(26), 3651–3664.

Watson, A. B., & Pelli, D. G. (1983). QUEST: A Bayesian adaptive psychometric method. *Perception and Psychophysics, 33*(2), 113–120.

Wurtz, R. H. (2008). Neuronal mechanisms of visual stability. *Vision Research, 48*(20), 2070–2089.

Zelinsky, G. J., & Neider, M. B. (2008). An eye movement analysis of multiple object tracking in a realistic environment. *Visual Cognition, 16*(5), 553–566.

Zelinsky, G., & Todor, A. (2010). The role of rescue saccades in tracking objects through occlusions. *Journal of Vision, 10*(7), 132 (abstract).