Visual Cognition

Visual working memory retains movement information within an allocentric reference frame

Justin N. Wood[a]

[a] Department of Psychology, University of Southern California, Los Angeles, CA, USA

First published on: 10 September 2010

PLEASE SCROLL DOWN FOR ARTICLE

Psychology Press
Taylor & Francis Group

Visual working memory retains movement information within an allocentric reference frame

Justin N. Wood

Department of Psychology, University of Southern California, Los Angeles, CA, USA

What frame of reference do we use to remember observed movements? One possibility is that visual working memory (VWM) retains movement information using a *retinotopic* frame of reference: A coordinate system with respect to the retina that retains view-dependent information. Alternatively, VWM might retain movement information using an *allocentric* frame of reference: A coordinate system with respect to the scene that retains view-invariant information. To address this question, I examined whether VWM retains view-dependent or view-invariant movement information. Results show that (1) observers have considerable difficulty remembering from which viewpoints they observed movements after a few seconds' delay, and (2) the same number of movements can be retained in VWM whether the movements are encoded and tested from the same viewpoint or from different viewpoints. Thus, movement representations contain little to no view-dependent information, which suggests that VWM uses an allocentric reference frame to retain movement information.

A primary function of the visual system is to track objects as they move. Whether the observer is a pedestrian tracking cars on the street or a wild animal tracking the movements of prey and predators, there is a need to maintain a visual index for objects that change in their spatial location over time.

Coordinated studies of adults, infants, and nonhuman animals provide evidence that object tracking depends on an innate,[1] evolutionarily ancient system (reviewed by Flombaum, Scholl, & Santos, 2009). In these studies, tracking abilities show four signature limits.[2] First, object tracking has a capacity limit: Adults, infants, and nonhuman animals can track a similar number of objects at once, with many experiments revealing a capacity limit of 3–4 objects (Barner, Wood, Hauser, & Carey, 2008; Feigenson, Carey, & Hauser, 2002; Hauser, Carey, & Hauser, 2000; Scholl, 2001), although object tracking performance also depends on the spacing between objects (Franconeri, Jonathan, & Scimeca, 2010). Second, object tracking privileges spatiotemporal information (where and when objects were encountered) over surface feature information (colour and shape). This privileged use of spatiotemporal information has been observed in several paradigms, including multiple object tracking (e.g., Scholl, Pylyshyn, & Franconeri, 1999), apparent motion (Dawson, 1991; Kolers, 1972), and the amodal integration of objects that move in and out of view (Burke, 1952; Flombaum, Kundey, Santos, & Scholl, 2004; Michotte, Thinès, & Crabbé, 1964/1991). Note, however, that surface feature information can influence object tracking in some contexts (e.g., Richard, Luck, & Hollingworth, 2008). Third, object tracking is sensitive to the manner in which objects disappear behind other objects. Adults and infants successfully track objects that disappear and reappear along fixed contours at the occluding boundaries, but fail to track objects that implode and explode into and out of existence at the occluding boundaries (Cheries, Feigenson, Scholl, & Carey, 2008; Scholl & Pylyshyn, 1999). Fourth, object tracking depends on an allocentric reference frame rather than a retinotopic reference frame. Adults and infants track objects equally well through 2-D and 3-D space (Kellman, Spelke, & Short, 1986; Liu, Austen, Booth, Fisher, Argue et al., 2005) and register the positions of objects in the environment rather than registering objects with respect to retinal locations (Fecteau, Chua, Franks, & Enns, 2001; Li & Warren, 2000; Liu, Healey, & Enns, 2003; von Hofsten, Kellman, & Putaansuu, 1992). These four signatures characterize an object tracking

---

[1] By *innate* I simply mean *not learned*. Learning mechanisms necessarily require unlearned abilities for detecting and analysing inputs and for drawing inferences, and so claims of learning inevitably presuppose a set of innate capacities (see Spelke, 1999; Spelke & Newport, 1998).

[2] The term "signature limits" refers to a consistent pattern of positive and negative findings that are obtained in tasks that require a particular set of cognitive processes (e.g., processes for tracking objects), and that are observed across studies that vary in methods, displays, and populations (e.g., infants, adults, and nonhuman animals). When such a pattern of findings is obtained, it can serve to test for the existence of those cognitive processes in further populations, for the situations that evoke them, and for the mechanisms that subserve them.

system that human adults appear to share with human infants and nonhuman primates, suggesting continuity in object tracking representations over human ontogeny and primate phylogeny.

Although there has been much interest in the origins and nature of the object-based attention mechanism that supports object tracking, much less is known about how we *retain* information about the movements of objects after movements have been observed. This is important because many cognitive and social tasks require the ability to remember movements, as opposed to simply tracking objects during immediate perception. For example, imitation requires tracking an individual's body movements, retaining information about those movements, and then later attempting to copy the movements based on information retained in memory. Thus, in order for movement information to guide behaviour, it must be stored in a temporary information buffer, known as visual working memory (VWM). But what is the nature of the VWM system that retains movement information?

This question can be broken into three more specific questions. First, what component of VWM retains movement information: Do observers remember movements with a VWM component that retains *spatial* information or with a more specialized VWM component that retains *spatiotemporal* information? To investigate this question, Wood (2007) used a dual-task method to measure observers' ability to remember movement information and static spatial information concurrently. In the first memory task observers attempted to remember varying numbers of observed movements, and in the second memory task observers attempted to remember varying numbers of object locations in a scene. Observers could remember the same number of movements whether they performed the first memory task alone or concurrently with the second memory task, and they could remember the same number of locations in a scene whether they performed the second memory task alone or concurrently with the first memory task. Thus, when observers performed two working memory tasks that required memory for movements and memory for the locations of objects in a scene, there was little to no competition between the two tasks for the limited storage resources of a single working memory system, despite both tasks requiring memory for spatial information. This finding provides evidence that VWM contains a specialized memory system for retaining observed movement information, which is distinct from the VWM system that retains information about the locations of objects in the scene (for convergent evidence see Smyth, Pearson, & Pendleton, 1989; Smyth & Pendleton, 1988). Additional dual-task experiments show that the VWM system that retains movement information can also be dissociated from the VWM systems that retain colour and shape information (Wood, 2008, 2010).

Second, what are the units of movement information retained in VWM: Are spatial and temporal features stored separately or as integrated

spatiotemporal representations? To address this question, Wood (2007) compared memory for movements defined by one feature (spatial or temporal feature) with memory for movements defined by a conjunction of features (spatial and temporal features). The results showed that it is possible to retain information about only 2–3 spatial or temporal features at one time. However, it is also possible to retain both the spatial and temporal features of 2–3 movements, indicating that VWM stores integrated spatiotemporal units rather than individual features.

Third, what spatial reference frame does this spatiotemporal-based VWM system use to represent observed movements: Do observers remember movements using a retinotopic, observer-centred reference frame or an allocentric, world-centred reference frame? The former hypothesis holds that VWM retains movement information in a *retinotopic* map: A coordinate system that represents locations with respect to the retina. Such retinotopic representations would contain viewpoint-specific information. The latter hypothesis holds that VWM retains movement information in an *allocentric* map: A coordinate system that represents locations with respect to the scene. Such allocentric representations would contain viewpoint-invariant information. The goal of the present study was to distinguish between these possibilities.

There is evidence that the visual system uses both retinotopic and allocentric reference frames to represent the visual world. One of the primary systems of place recognition in humans and nonhuman animals is a view-dependent "snapshot" system that stores retinotopic representations. In brief, animals take a visual "snapshot" of the scene surrounding a target goal and store this view in memory. During navigation, the animal moves in order to recover this target view so as to reduce the difference between the current view and the target view (see Collett & Collett, 2000).

Evidence for snapshot representations comes from studies of navigating insects and mammals. Bees, for example, were trained to forage in an environment filled with landmarks and then the locations of the food source and the landmarks were moved. Bees approached the food source from a constant direction, so that the visual image of the scene was roughly the same each time they approached the food (Collett & Lehrer, 1993; Collett & Rees, 1997). Some insects such as wood ants store multiple snapshots of a familiar landmark from different vantage points so that they may approach a familiar landmark from multiple angles (Judd & Collett, 1998). Snapshot representations also guide visual navigation in rodents and humans. For instance, rodents in a water maze tend to approach a hidden support from a particular direction (Sutherland, Chew, Baker, & Linggard, 1987). Similarly, human adults readily learn locations in virtual reality environments defined entirely by a continuous colour gradient without individual objects that may be used as landmarks, in qualitative agreement with a view-dependent snapshot

system but not with other models of place recognition (Gillner, Weib, & Mallot, 2008).

In contrast to these retinotopic representations, other visual processes use an allocentric coordinate system to represent visual information. For instance, spatial working memory uses an allocentric reference frame to retain information about locations (e.g., Jiang, Olson, & Chun, 2000; Phillips, 1974). Similarly, the object tracking system uses an allocentric frame of reference to track the locations of objects over time. At least four types of evidence support this conclusion. First, observers track objects equally well whether the objects move in 2-D or 3-D space (Liu et al., 2005). Second, reductions in scene coherence (e.g., distorting the perception of 3-D space) impair tracking accuracy (Liu et al., 2005). Third, saccades that are made from one location to another are referenced to stationary environmental landmarks rather than to specific retinal coordinates (Deubel, Bridgeman, & Schneider, 1998). Fourth, smooth pursuit eye movements are linked to retinal rather than to environmental coordinates (Raymond, Shapiro, & Rose, 1984). Together, these findings suggest that visual perception registers the positions of objects in the environment rather than registering objects with respect to their retinal locations.

The present study examines whether the spatiotemporal-based VWM system retains information using retinotopic coordinates, akin to the view-dependent snapshot system, or allocentric coordinates, akin to the object tracking system. If movement information is retained using a retinotopic frame of reference, then the movement information retained in VWM will be view-dependent because retinotopic representations contain viewpoint-specific features. Thus, viewpoint information would be an integral part of the movement representations retained in VWM. Alternatively, if movement information is retained using an allocentric frame of reference, then the movement information retained in VWM will be viewpoint-invariant because allocentric representations register locations with respect to environmental rather than retinal coordinates. In consequence, a movement representation in VWM would not contain viewpoint-specific features.

To distinguish between these possibilities, I tested two unique predictions of a memory mechanism that uses an allocentric frame of reference. First, observers should have difficulty remembering from which viewpoints they observed movements, even after only a few seconds' delay. A movement representation that is based on allocentric coordinates will not contain viewpoint-specific information. Binding movement and viewpoint information would therefore require additional VWM resources, or it might not happen at all. In contrast, if movement representations are based on a retinotopic reference frame, then observers should have no difficulty remembering from which viewpoints they observed movements because retinotopic representations contain viewpoint-specific information.

Second, if movement representations are based on an allocentric reference frame, then observers should be able to remember the same number of movements whether the movements are encoded and tested from the same viewpoint or from different viewpoints. Changing the viewpoint between the study and test movements should have relatively little effect on an allocentric memory mechanism because the movements would be registered with respect to environmental rather than retinal coordinates, and would therefore be viewpoint-invariant. However, if movement representations are based on a retinotopic reference frame, then changing the viewpoint of the test movement should severely impair memory performance because the features of the retinotopic representations stored in VWM would not match with the visible features of the test movement even when the movements were the same (see Figure 4).

Experiment 1 tested the first prediction, by examining whether observers have difficulty remembering from which viewpoints they observed movements after a few seconds' delay. Experiment 2 tested the second prediction, by examining whether observers can remember similar numbers of movements whether the movements are encoded and tested from the same viewpoint or from different viewpoints. To preview the findings, the results confirmed both predictions of a VWM mechanism that uses an allocentric reference frame.

EXPERIMENT 1

To examine whether observers have difficulty remembering from which viewpoints they observed movements after a few seconds' delay, I used the sequential comparison procedure used previously to study VWM for observed movements (Wood, 2007, 2008) and objects (e.g., Luck & Vogel, 1997). On each trial, participants viewed a study sequence in which a computer-animated figure performed three different movements, each of which was observed from a different viewpoint (see Figure 1). After a brief delay interval, the figure performed a test movement, and participants indicated whether that test movement was the same or different from the study movements with respect to the features detailed below. During each trial, participants performed an articulatory suppression task to prevent verbal recoding of the stimuli (Besner, Davies, & Daniels, 1981).

Participants completed four different conditions. In the *movements only* condition, participants needed to remember the movements only. This condition provided a baseline measure of the number of movements that can be retained in VWM at once. In the *viewpoints only* condition, participants needed to remember only the viewpoints from which they observed the movements. This condition provided a baseline measure of the number of
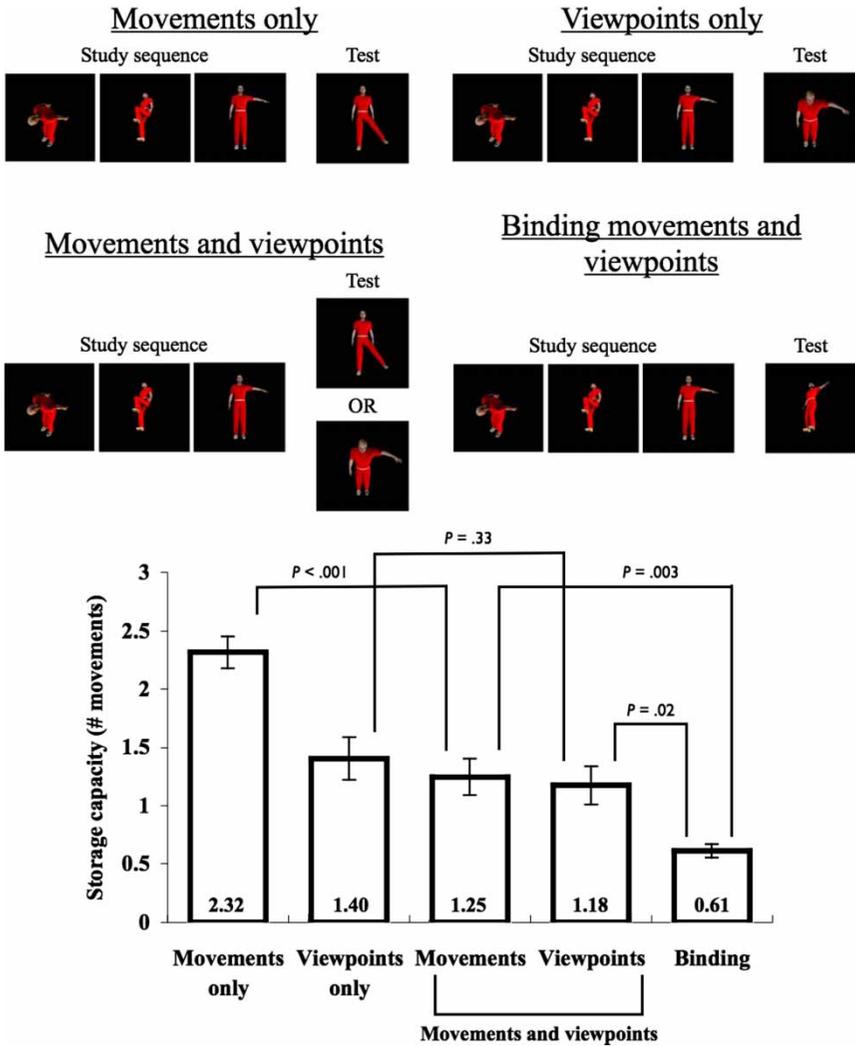
Figure 1. Sample stimuli and results from Experiment 1. (Top) Schematic illustration of a trial from each of the four conditions. All examples depict different trials. (Bottom) Memory capacity estimates from each condition, as well as *p*-values denoting the statistical differences between conditions and trial types as computed through paired-samples *t*-tests. Error bars denote standard error. To view this figure in colour, please see the online issue of the Journal.

viewpoints that can be retained in VWM at once. In the *movements and viewpoints* condition, participants needed to remember the movements and the viewpoints simultaneously. This condition provided a measure of the number of movements and viewpoints that can be retained in VWM

concurrently. In the *binding movements and viewpoints* condition, participants needed to remember the movements and viewpoints in an integrated form. This condition provided a measure of the number of integrated movement-viewpoint representations that can be retained in VWM at once.

If movement information is retained within a retinotopic frame of reference, then performance should be similar across the four conditions because the movement and viewpoint information from the study sequences would be stored together, within integrated representations. Thus, the four conditions would place equivalent demands on VWM. However, if movement information is retained using an allocentric frame of reference, then performance will differ across the conditions. In particular, two specific patterns should emerge. First, performance should be lower in the movements and viewpoints condition compared to in the movements only condition because the movement representations stored in VWM would not contain viewpoint-specific information. Thus, additional VWM resources would be needed to remember the viewpoints from which movements were observed. Second, performance should be lower in the binding movements and viewpoints condition compared to in the movements and viewpoints condition because movement and viewpoint information are not automatically bound together within a memory mechanism that uses an allocentric frame of reference. Thus, additional VWM resources would be needed to bind movement and viewpoint information into an integrated representation.

Method

Participants.   Ten participants (two males, eight females) between the ages of 18 and 30 ($M = 23.2$, $SD = 4.39$) with normal or corrected-to-normal vision participated to receive credit towards a course requirement or for monetary payment. Informed consent was obtained.

Design.   On each trial, participants viewed a study sequence consisting of three different movements, each of which was observed from a different viewpoint. After a brief delay interval, a test movement was observed, and participants indicated whether the relevant features of that test movement had been present in the study sequence, with respect to the following conditions:

1. *Movements only*: Participants were told that only the movements could change and to remember only the movements. On different trials, the test movement was different from all three of the movements in the study sequence. The test movement was observed from a viewpoint that had been associated with one of the three movements from the study sequence.

2. *Viewpoints only*: Participants were told that only the viewpoints could change and to remember only the viewpoints. On different trials, the test movement (one of the three movements from the study sequence) was observed from a viewpoint that none of the movements in the study sequence had been observed from.
3. *Movements and viewpoints*: Participants were told that either the movements or the viewpoints could change and to remember both the movements and the viewpoints. On 50% of the different trials, the test display presented a movement that had not been present in the study sequence. On the other 50% of the different trials, the test movement was observed from a viewpoint that none of the movements in the study sequence had been observed from.
4. *Binding movements and viewpoints*: The test display always consisted of a movement and a viewpoint that had been present in the study sequence. However, on different trials, the test movement was observed from a viewpoint that one of the other movements in the study sequence had previously been observed from. Participants were told to treat such changes as different.

Participants received 50 trials in each condition. Each condition was preceded by six practice trials. The order of conditions was counterbalanced across participants in a within-subjects design.

Procedure. Each trial began with a 1000 ms presentation of two randomly selected letters, and participants were required to repeat those letters continuously and out loud until the end of the trial. The offset of these letters was followed by a 1000 ms presentation of a screen displaying the word "ready", followed by the presentation of the study sequence. The study sequence consisted of a computer-animated figure performing three movements, each of which was observed from a different viewpoint. All movements were dynamic (i.e., they involved fluid, continuous movement, rather than being presented as static pictures). The animations were created using Poser 6 software from SmithMicro. From the front viewpoint, the figure subtended $10.5°$ (height)$\times 4°$ (width) in the centre of a video monitor. Each movement lasted 500 ms and was followed by 500 ms of stasis. Then, the figure disappeared for 200 ms and reappeared from a new viewpoint. The movements were selected at random without replacement from a set of seven highly discriminable movements: Forearm curl, arm raise, head turn, body twist, knee raise, leg raise, and torso bend (see Figure 2). The figure performed the movements on the left side of his body. The viewpoints were selected at random without replacement from a set of nine highly discriminable viewpoints (see Figure 2).
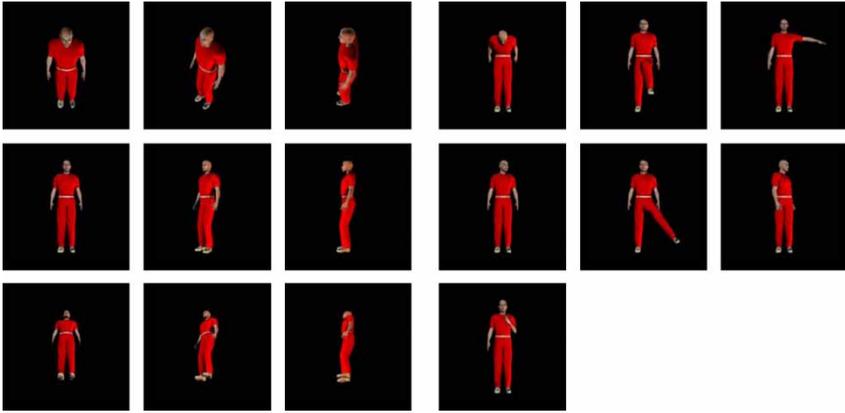
Figure 2.   Static depictions of the stimuli used in Experiment 1. (Left) Images showing the nine different viewpoints. (Right) Images showing the seven movements by illustrating the movements' maximal deviation from the neutral position shown in the images on the left. To view this figure in colour, please see the online issue of the Journal.

The study sequence was followed, after a 250 ms delay, by an 800 ms presentation of the word "test", followed by the presentation of the test display, which consisted of the figure performing a single movement from one of the nine viewpoints. Participants were required to make a response to the test display, indicating whether the relevant features of the test display had been present in the study sequence.

Results

For the statistical analyses, the data were converted into capacity estimates by using the formula developed by Cowan (2001). The logic of this approach is that if an observer can retain k items from a sequence consisting of n items, then the observer should be able to detect a change in one of the items on k/n trials. This approach takes into consideration the effects of guessing, by factoring in the false alarm rate (F = false alarms/(false alarms + correct rejections) and the observed hit rate (H = hits/(hits + misses). The formula is defined as $k = n$ ($H - F$). In the *movements and viewpoints* condition, the average F-value of the "same" trials was used for the separate statistical analyses of the movement trials and the viewpoint trials. For all experiments, the same statistical patterns were observed when accuracy was used as the dependent measure.

A repeated measures ANOVA revealed a main effect of condition, $F(3, 27) = 31.87$, $p < .001$, $h_p^2 = .78$. Post hoc analyses revealed the pattern

of results shown in Figure 1 (see Appendix 1 for the hits, false alarms, and reaction times for all conditions). Memory capacity in the movements only condition (2.32 movements) was significantly higher than in the viewpoints only condition (1.40 viewpoints). Memory capacity in the movements and viewpoints condition for the movement information (1.25 movements) was significantly lower than memory capacity in the movements only conditions, $t(9) = 7.27$, $p < .001$. Further, memory capacity was significantly lower in the binding movements and viewpoints condition (0.61 integrated representations), in which participants needed to remember which movement was observed from which viewpoint, than in the movements and viewpoints condition, in which participants needed to remember the movements and the viewpoints but not in an integrated form, $t(9) = 3.78$, $p < .005$.

Discussion

These results provide evidence that VWM representations observed movements are view-invariant and thus stored with respect to an allocentric reference frame. Two specific patterns support this conclusion. First, memory capacity was lower when participants needed to remember movements and viewpoints concurrently compared to when they needed to remember movements and viewpoints seperately. Second, memory capacity was lower when participants needed to retain integrated movement-viewpoint representations compared to when they needed to retain movement and viewpoint representations concurrently but not in an integrated form. Neither of these patterns should have emerged if movements were retained with respect to a retinotopic frame of reference because retinotopic representations contain viewpoint-specific information.

There are two alternative explanations for this pattern of data. First, participants may have retained view-dependent representations but encoded different low-level features of the stimuli in the movements and viewpoints condition compared to in the movements only condition. Thus, on different trials, the degree to which the test stimuli differed from the view-dependent representations retained in VWM may have been smaller in the movements and viewpoints condition compared to in the movements only condition, thereby leading to lower performance. Similarly, participants may have retained view-dependent representations but encoded different low-level features of the stimuli in the binding movements and viewpoints condition compared to in the movements and viewpoints condition. Thus, on different trials, the degree to which the test stimuli differed from the view-dependent representations retained in VWM may have been smaller in the binding movements and viewpoints condition compared to in the movements and viewpoints condition, thereby leading to lower performance. Unfortunately,

this alternative explanation is difficult to test. Unlike studies that measure VWM for simple 2-D stimuli (e.g., coloured bars presented at different orientations), which allow precise control over the magnitude of the change between a stored representation and a test stimulus (e.g., the change from red to green with oriented bars is psychophysically identical whether the object is horizontal or vertical), it is not possible to so precisely control change magnitude with the 3-D stimuli used in the present study.

Second, in the movements and viewpoints condition both the movements and the viewpoints could change. Thus, it is impossible to know whether the false alarms in this condition reflected incorrect perceptions of movement changes or incorrect perceptions of viewpoint changes. Because the viewpoints only condition was more difficult than the movements only condition, it is possible that the majority of the false alarms in the movements and viewpoints condition arose because of incorrect perceptions of viewpoint changes. The large drop in storage capacity for the movement information in the movements and viewpoints condition compared to the movements only condition may therefore have been an artifact.

To distinguish between these alternative accounts, Experiment 2 tested a different prediction of a memory mechanism that uses an allocentric frame of reference: Observers should be able to remember the same number of movements whether the movements are encoded and tested from the same viewpoint or from different viewpoints.

To test this prediction, observers were asked to remember three study movements, and then they were shown a test movement from the same viewpoint or from viewpoints that differed 45° or 90° from the study viewpoint. Observers then indicated whether the test movement was one of the three study movements. If VWM uses a retinotopic frame of reference, then changing the viewpoint of the test movement should severely impair memory performance because the view-dependent features of the representations stored in VWM would not match with the visible features of the test movement even when the movements were the same. However, if VWM uses an allocentric frame of reference to retain movement information, then changing the viewpoint between the study and test movements should have no effect on performance because movements would be retained within a reference frame that registers locations with respect to the environment rather than with respect to retinal locations.

EXPERIMENT 2

Participants observed three study movements and a test movement, and then indicated whether the test movement was one of the three study movements.

On one set of trials, the test movement was presented from the same viewpoint as the study movements. On a second set of trials, the test movement was presented from a viewpoint that differed by 45° from the study viewpoint. On a third set of trials, the test movement was presented from a viewpoint that differed by 90° from the study viewpoint (see Figure 3). If observers retain view-dependent representations within a retinotopic frame of reference, then performance should decrease as the difference between the vantage points of the study and test movements increases. Thus, performance should be best when the study and test movements are observed from the same viewpoint, lower when the test movement viewpoint differs by 45° from the study movement viewpoint, and lowest when the test movement viewpoint differs by 90° from the study movement viewpoint.

Any potential viewpoint-dependent effects that arise in this experiment can not be due to certain viewpoints being inherently easier or harder to process because all pairwise combinations of viewpoints were tested. That is, the 0° viewpoint difference condition included trials testing all three of the 0°, 45°, and 90° viewpoints, the 45° viewpoint difference condition included trials testing 0°–45° viewpoint combinations and 45°–90° viewpoint combinations in both possible study–test orders, and the 90° viewpoint difference condition included 0°–90° trials and 90°–0° trials. Thus, any decrease in performance from 0° to 45° to 90° must result from the changes of viewpoint in the latter two conditions, not from the particular viewpoints that were tested.

Method

Participants.   Ten new participants (three males, seven females) between the ages of 18 and 33 ($M = 21.9$, $SD = 4.33$) participated.

Procedure.   Each trial began with a 1000 ms presentation of two randomly selected letters, and participants were required to repeat those letters continuously and out loud until the end of the trial. The offset of these letters was followed by a 500 ms presentation of a black screen followed by the presentation of the study sequence. The study sequence consisted of a computer-animated figure performing three movements. Each movement lasted 500 ms and was followed by 500 ms of stasis. The movements were selected at random without replacement from a set of seven highly discriminable movements: Arm raise, body twist, forearm curl, knee raise, leg raise (to the front of the body), leg raise (to the side of the
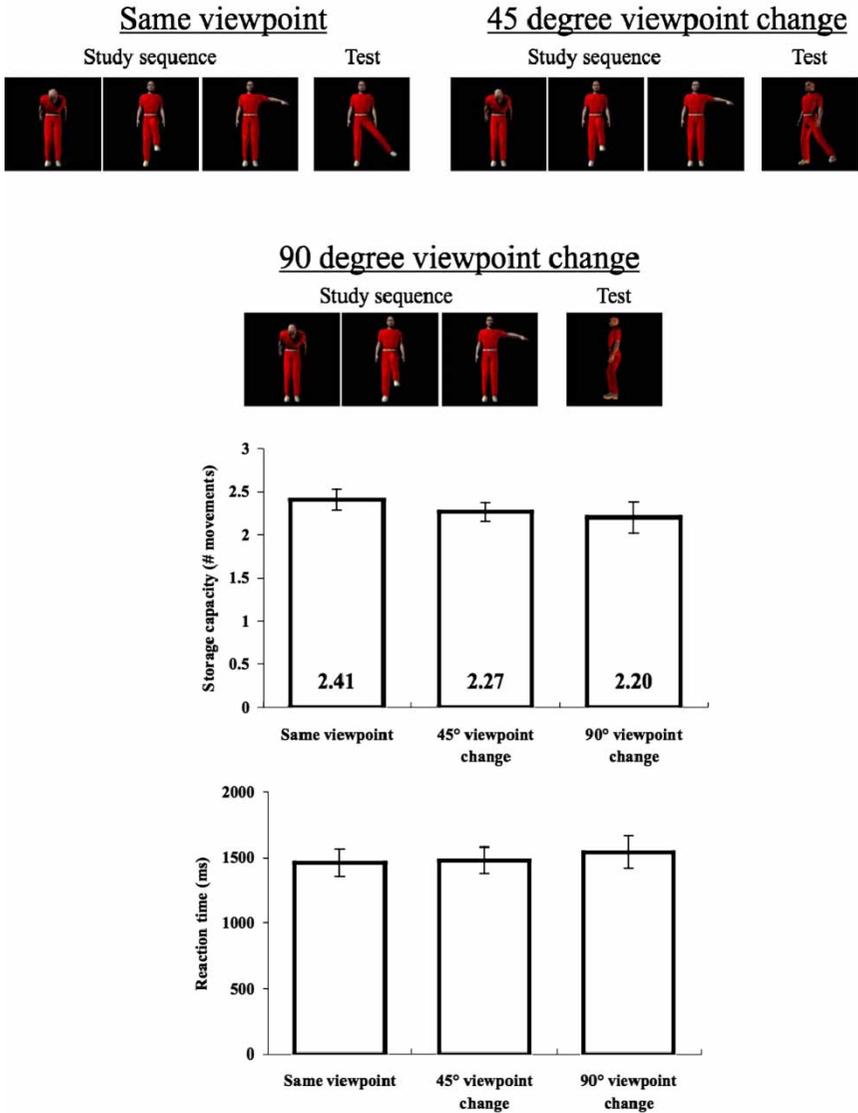
Figure 3.   Sample stimuli and results from Experiment 2. (Top) Schematic illustration of a trial. (Middle) Memory capacity estimates when the study and test viewpoints differed by 0°, 45°, and 90°. (Bottom) Reaction times to the test movement when the study and test viewpoints differed by 0°, 45°, and 90°. Error bars denote standard error. To view this figure in colour, please see the online issue of the Journal.
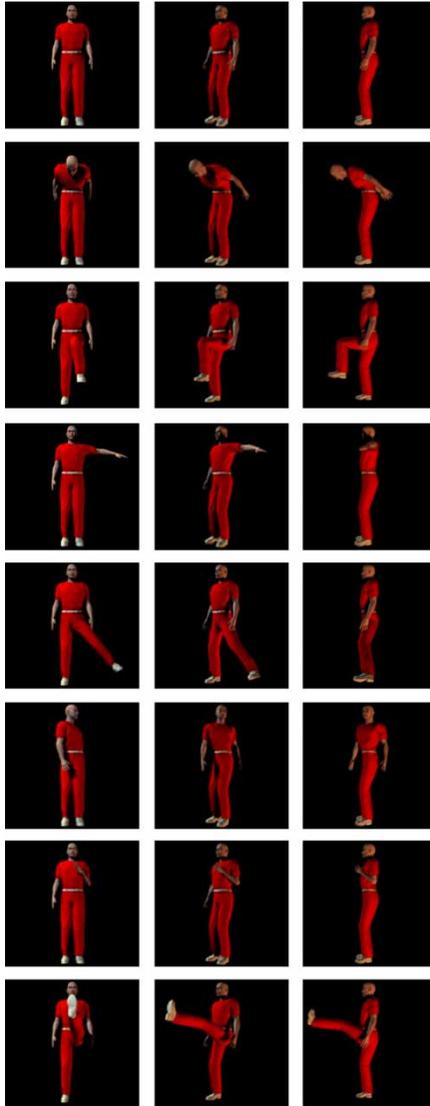
Figure 4. Static depictions of the stimuli used in Experiment 2. The images show the movements' maximal deviation from the neutral position shown at the top of the figure. All movements are shown from the three different viewpoints. To view this figure in colour, please see the online issue of the Journal.

body), and torso bend (see Figure 4). All of the study movements in a trial were observed from the same viewpoint.

The study sequence was followed by a 1000 ms delay interval consisting of a black screen and then the test movement. Participants made a response to

the test movement, indicating whether the movement had been present in the study sequence. Participants were not instructed to respond quickly; nevertheless, reaction times were collected to assess potential speed–accuracy tradeoffs.

The viewpoint from which the study movements and test movement were observed varied systematically. The study and test movements were observed from the front view, a 45° horizontal, orbital rotation from the front view, and a 90° horizontal, orbital rotation from the front view (see Figure 4 for static depictions of all movements from each viewpoint). Participants received 270 trials, which were preceded by eight practice trials. Participants received 30 trials for each of the nine unique pairwise combination of these viewpoints across the study and test movements: 0°–0°, 0°–45°, 0°–90°, 45°–0°, 45°–45°, 45°–90°, 90°–0°, 90°–45°, & 90°–90°.

Results and discussion

For the analyses, the data from the nine pairwise combinations described previously were divided into three comparison groups as follows: 0° difference between study and test viewpoints (0°–0°, 45°–45°, 90°–90°), 45° difference between study and test viewpoints (0°–45°, 45°–0°, 45°–90°, 90°–45°), and 90° difference between study and test viewpoints (0°–90°, 90°–0°).

Figure 3 presents the results (see Appendix 2 for the hits, false alarms, and reaction times for all pairwise combinations of study and test viewpoints). A repeated measures ANOVA did not reveal a significant difference in storage capacity when the study viewpoint and the test viewpoint differed by 0°, 45°, or 90°, $F(2, 18) = 1.72$, $p = .21$, $h_p^2 = .16$. On average, observers retained 2.41 movements when the study and test viewpoints differed by 0°, 2.27 movements when the study and test viewpoints differed by 45°, and 2.20 movements when the study and test viewpoints differed by 90°. Furthermore, a repeated measures ANOVA did not reveal a significant difference in reaction time when the study viewpoint and the test viewpoint differed by 0°, 45°, or 90°, $F(2, 18) = 1.12$, $p = .35$, $h_p^2 = .11$. Thus, the number of movements that participants could remember was similar whether the study and test viewpoints differed by 0°, 45°, or 90°, and the speed at which they could indicate whether a test movement was the same or different from one of the study movements was similar whether the study and test viewpoints differed by 0°, 45°, or 90°. This pattern of results suggests that movement representations retained in VWM contain little to no viewpoint-specific information. These results converge with the results of Experiment 1, providing additional evidence that VWM retains view-invariant movement representations within an allocentric reference frame.

GENERAL DISCUSSION

Previous studies provide evidence for a specialized visual working memory (VWM) system that retains integrated spatiotemporal representations of observed movements (Wood, 2007, 2008, 2010). The goal of the present study was to characterize the type of spatial reference frame used by this VWM system. One possibility is that VWM retains movement information using a *retinotopic*, observer-centred frame of reference: A coordinate system with respect to the retina that retains view-dependent information. Alternatively, VWM might retain movement information using an *allocentric*, world-centred frame of reference: A coordinate system with respect to the scene that retains view-invariant information. To distinguish between these possibilities, I tested two unique predictions of a memory mechanism that stores information within an allocentric frame of reference. First, observers should have difficulty remembering from which viewpoints they observed movements after a few seconds' delay. Second, observers should be able to remember the same number of movements whether the movements are encoded and tested from the same viewpoint or from different viewpoints. These predictions were confirmed in Experiments 1 and 2, respectively, which suggests that VWM retains movement information within an allocentric reference frame.

In Experiment 1, after observing a sequence of movements, participants had difficulty remembering which movements were observed from which viewpoints. The surprisingly low performance in the binding movements and viewpoints condition suggests that even when retaining information about movements and viewpoints, participants had a difficult time binding these two types of information together. On average, participants retained only 0.61 integrated movement/viewpoint representations, meaning that they successfully represented a single integrated representation on just a subset of the trials. In contrast, when participants did not need to remember integrated movement/viewpoint representations, they were able to retain information about 1–2 movements and 1–2 viewpoints on each trial. Thus, binding movement and viewpoint information into an integrated representation in working memory appears to require additional resources over and above those used to retain movement and viewpoint information in a nonintegrated form.

In Experiment 2, participants attempted to remember three study movements observed from one viewpoint, and then compared those movements to a test movement observed from the same viewpoint or from a viewpoint that differed by 45° or 90° from the study viewpoint. On average, participants retained 2.41 movements when the study and test viewpoints were identical, 2.27 movements when the study and test viewpoints differed from one another by 45°, and 2.20 movements when the study and test

viewpoints differed from one another by 90°. Thus, nearly the same number of movements could be retained in VWM whether the movements were encoded and tested from the same viewpoint or from different viewpoints. Furthermore, participants' reaction times were similar whether the test movement was presented from the same viewpoint or from a different viewpoint as the study movements. Movement representations in VWM therefore appear to contain little to no view-dependent information. This experiment provides further evidence that VWM retains movement information within an allocentric reference frame.

These results place constraints on the cognitive and neurobiological mechanisms subserving the VWM system that retains spatiotemporal information. In particular, VWM tasks have been found to depend on neural substrates from both early (V1–V4) and late (e.g., lateral occipital cortex) levels of the visual hierarchy (e.g., Harrison & Tong, 2009; Vogel & Machizawa, 2004). In principle, either of these substrates could retain movement information. However, representations in early levels of the hierarchy are pixel-like pictorial representations of the scene, akin to photographs. VWM representations retained within these early levels of the hierarchy therefore contain view-specific image features. In contrast, later levels of the hierarchy support object and movement recognition mechanisms that represent the perceived 3-D shape and path of motion of objects (e.g., Kable & Chatterjee, 2006; Kourtzi & Kanwisher, 2001), but do not represent view-specific image features. VWM representations retained within these later levels of the hierarchy may therefore be view invariant. Thus, the view-invariant movement representations identified in the present study are most likely retained in later levels of the visual hierarchy, rather than in the earlier levels, which retain view-specific image features.

These results leave open the possibility that VWM can retain a small amount of movement information within view-dependent representations. Specifically, in Experiment 1 participants retained roughly 0.61 movement's worth of information within view-specific representations on each trial, and in Experiment 2 there was a nonsignificant cost of 0.21 movement's worth of information when the movements were encoded and tested from viewpoints that differed by 90° compared to 0°. It will be interesting for future studies to examine whether such view-dependent representations are retained in the same VWM system that retains view-invariant movement representations, or whether view-dependent and view-invariant movement representations are supported by separate, specialized memory system. For example, with regard to object representation, there is some evidence that VWM contains separate, specialized memory systems for retaining view-dependent "snapshot" information and view-invariant object identity information (Wood, 2009, 2010).

Linking visual working memory to the biological foundations of cognition

As discussed in the introduction, coordinated studies of adults, infants, and nonhuman animals provide evidence for an innate, evolutionarily ancient object tracking system that operates primarily over spatiotemporal representations. Similarly, studies of VWM provide evidence for a specialized memory system that retains integrated spatiotemporal representations (Wood, 2007, 2008, 2010). The present study builds on these findings by showing that VWM uses an allocentric frame of reference to represent movement information, as does the object tracking system (Liu et al., 2005). Thus, the spatiotemporal-based VWM system and the object tracking system share two information-processing signatures. First, both mechanisms use the same basic elements—integrated spatiotemporal units— to represent the movements of objects. Second, both mechanisms use an allocentric spatial reference frame to represent those spatiotemporal units. These shared information-processing signatures suggest that the spatiotemporal-based VWM system and object tracking system depend on common neural mechanisms. This is important because it links the study of VWM to the study of the biological foundations of visual cognition (Wood, 2010).

This link may be fruitful in many ways. For the study of visual cognition, these findings link the spatiotemporal-based VWM system to the innate, evolutionarily ancient object tracking system that strutures visual experience from birth. This could illuminate the origins of the spatiotemporal-based VWM system, on both developmental and evolutionary timescales. For the study of the biological foundations of cognition, these findings show how our innate, evolutionarily ancient object tracking system corresponds to one of the VWM systems that supports and limits mature visual thought.

REFERENCES

Barner, D., Wood, J. N., Hauser, M. D., & Carey, S. (2008). Wild rhesus monkeys compute the singular-plural distinction. *Cognition*, *107*, 603–622.

Besner, D., Davies, J., & Daniels, S. (1981). Reading for meaning: The effects of concurrent articulation. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *33A*, 415–437.

Burke, L. (1952). On the tunnel effect. *Quarterly Journal of Experimental Psychology*, *4*, 121–138.

Cheries E., Feigenson, L., Scholl, B. J., & Carey, S. (2008). Cues *to object persistence in infancy: Tracking objects through occlusion vs. implosion*. Poster presented at the International Conference on Infant Studies. 3/29/08, Vancouver, BC, Canada.

Collett, M., & Collett, T. S. (2000). How do insects use path integration for their navigation? *Biological Cybernetics*, *83*(3), 245–259.

Collett, T. S., & Lehrer, M. (1993). Looking and learning: A spatial pattern in the orientation flight of the wasp *Vespula vulgaris*. *Proceedings of the Royal Society of London*, *252B*, 129–134.

Collett, T. S., & Rees, J. A. (1997). View-based navigation in Hymenoptera: Multiple strategies of landmark guidance in approach to a feeder. *Journal of Comparative Physiology*, *181A*, 47–58.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*, 87–185.

Dawson, M. (1991). The how and why of what went where in apparent motion: Modeling solutions to the motion correspondence problem. *Psychological Review*, *98*, 569–603.

Deubel, H., Bridgeman, B., & Schneider, W. X. (1998). Immediate post-saccadic information mediates space constancy. *Vision Research*, *38*, 3147–3159.

Fecteau, J. H., Chua, R., Franks, I., & Enns, J. T. (2001). Visual awareness and the on-line modification of action. *Canadian Journal of Experimental Psychology*, *55*, 104–110.

Feigenson, L., Carey, S., & Hauser, M. (2002). The representations underlying infants' choice of more: Object-files versus analog magnitudes. *Psychological Science*, *13*, 150–156.

Flombaum, J. I., Kundey, S., Santos, L. R., & Scholl, B. J. (2004). Dynamic object individuation in rhesus macaques: A study of the tunnel effect. *Psychological Science*, *15*, 795–800.

Flombaum, J. I., Scholl, B. J., & Santos, L. R. (2009). Spatiotemporal priority as a fundamental principle of object persistence. In B. Hood & L. Santos (Eds.), *The origins of object knowledge* (pp. 135–164). Oxford, UK: Oxford University Press.

Franconeri, S. V., Jonathan, S. V., & Scimeca, J. M. (2010). Tracking multiple objects is limited only by object spacing, not by speed, time, or capacity. *Psychological Science*. Advance online publicaion. doi: 10.1177/0956797610373935.

Gillner, S., Weiß, A. M., & Mallot, H. A. (2008). Visual homing in the absence of feature-based landmark information. *Cognition*, *109*, 105–122.

Harrison, S. A., & Tong, F. (2009). Decoding reveals the contents of visual working memory in early visual areas. *Nature*. doi: 10.1038/nature07832.

Hauser, M. D., Carey, S., & Hauser, L. B. (2000). Spontaneous number representation in semi-free-ranging rhesus monkeys. *Proceedings of the Royal Society of London*, *267B*, 829–833.

Jiang, Y., Olson, I. R., & Chun, M. M. (2000). Organization of visual short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 683–702.

Judd, S. P. D., & Collett, T. S. (1998). Multiple stored views and landmark guidance in ants. *Nature*, *392*, 710–714.

Kable, J. W., & Chatterjee, A. (2006). Specificity of action representations in the lateral occipital cortex. *Journal of Cognitive Neuroscience*, *18*(9), 1498–1517.

Kellman, P. J., Spelke, E. S., & Short, K. (1986). Infant perception of object unity from translatory motion in depth and vertical translation. *Child Development*, *57*, 72–86.

Kolers, E. (1972). *Aspects of motion perception*. New York: Pergamon Press.

Kourtzi, Z., & Kanwisher, N. (2001). Representation of perceived object shape by the human lateral occipital complex. *Science*, *293*, 1506–1509.

Li, L., & Warren, W. H. (2000). Perception of heading during rotation: Sufficiency of dense motion parallax and reference objects. *Vision Research*, *40*, 3873–3894.

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.

Liu, G., Healey, C. G., & Enns, J. T. (2003). Target detection and localization in visual search: A dual systems perspective. *Perception and Psychophysics*, *65*, 678–694.

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*, 279–281.

Michotte, A., Thinès, G., & Crabbé, G. (1991). Amodal completion of perceptual structures. In G. Thines, A. Costall, & G. Butterworth, (Eds.), *Michotte's experimental phenomenology of*

perception (pp. 140–167). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc. (Reprinted and translated from Les complements amodaux des structures perceptives, in *Studia Psychologica*, by G. Thines, A. Costall, & G. Butterworth, Eds., 1991, Louvain, Belgium: Publications Universitaires)

Phillips, W. A. (1974). On the distinction between sensory storage and short-term visual memory. *Perception and Psychophysics*, *16*, 283–290.

Raymond, J. E., Shapiro, K. L., & Rose, D. J. (1984). Optokinetic backgrounds affect perceived velocity during ocular tracking. *Perception and Psychophysics*, *36*, 221–224.

Richard, A. M., Luck, S. J., & Hollingworth, A. (2008). Establishing object correspondence across eye movements: Flexible use of spatiotemporal and surface feature information. *Cognition, 109*, 66–88

Scholl, B. J. (2001). Objects and attention: The state of the art. *Cognition*, *80*, 1–46.

Scholl, B. J., & Pylyshyn, Z. W. (1999). Tracking multiple items through occlusion: Clues to visual objecthood. *Cognitive Psychology*, *38*, 259–290.

Scholl, B. J., Pylyshyn, Z. W., & Franconeri, S. (1999). When are spatiotemporal and featural properties encoded as a result of attentional allocation?. *Investigative Ophthalmology and Visual Science*, *40*(4), S797.

Smyth, M. M., Pearson, N. A., & Pendleton, L. R. (1989). Movement and working memory: Patterns and position in space. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *40A*, 497–514.

Smyth, M. M., & Pendleton, L. R. (1988). Working memory for movements. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *41A*, 235–250.

Spelke, E. S. (1999). Innateness, learning, and the development of object representation. *Developmental Science*, *2*, 145–148.

Spelke, E. S., & Newport, E. (1998). Nativism, empiricism, and the development of knowledge. In R. Lerner (Ed.), *Handbook of child psychology: Vol. 1. Theoretical models of human development* (5th ed.) pp. 275–340. New York: Wiley.

Sutherland, R. J., Chew, G. L., Baker, J. C., & Linggard, R. C. (1987). Some limitations on the use of distal cues in place navigation by rats. *Psychobiology*, *15*, 48–57.

Vogel, E. K., & Machizawa, M. G. (2004). Neural activity predicts individual differences in visual working memory capacity. *Nature*, *428*, 748–751.

Von Hofsten, C., Kellman, P. J., & Putaansuu, J. (1992). Young infants' sensitivity to motion parallax. *Infant Behavior and Development*, *15*, 245–264.

Wood, J. N. (2007). Visual working memory for observed actions. *Journal of Experimental Psychology: General*, *136*(4), 639–652.

Wood, J. N. (2008). Visual memory for agents and their actions. *Cognition*, *108*, 522–532.

Wood, J. N. (2009). Distinct visual working memory systems for view-dependent and view-invariant representation. *PLoS ONE*, *4*(8), e6601. doi: 10.1371/journal.pone.0006601.

Wood, J. N. (2010). A core knowledge architecture of visual working memory. Manuscript submitted for publication.

APPENDIX 1

The proportion of hits (responding *different* on change trials), the proportion of false alarms (responding *different* on same trials), and the reaction times for each of the conditions in Experiment 1

Condition	Hits/FAs	RT (ms)
Movements only	.87/.10	1654
Viewpoints only	.71/.24	1582
Movements & viewpoints		
Movements	.82/.40	1888
Viewpoints	.79/.40	1949
Binding	.60/.39	1838

APPENDIX 2

The proportion of hits (responding *different* on change trials), the proportion of false alarms (responding *different* on same trials), and the reaction times for each of the conditions in Experiment 2

Condition	Hits/FAs	RT (ms)
0–0	.90/.11	1480
0–45	.86/.11	1446
0–90	.84/.09	1535
45–0	.82/.09	1453
45–45	.87/.08	1441
45–90	.87/.09	1489
90–0	.83/.11	1549
90–45	.93/.17	1524
90–90	.86/.03	1460