The Quarterly Journal of Experimental Psychology
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/pqje20

Posture-based processing in visual short-term memory for actions

Staci A. Vicary[a] & Catherine J. Stevens[ab]

[a] The MARCS Institute, University of Western Sydney, Penrith, NSW, Australia

[b] School of Social Sciences and Psychology, University of Western Sydney, Penrith, NSW, Australia

Accepted author version posted online: 24 Jun 2014.Published online: 10 Jul 2014.

To cite this article: Staci A. Vicary & Catherine J. Stevens (2014) Posture-based processing in visual short-term memory for actions, The Quarterly Journal of Experimental Psychology, 67:12, 2409-2424, DOI: 10.1080/17470218.2014.931445

To link to this article: http://dx.doi.org/10.1080/17470218.2014.931445

THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY, 2014
Vol. 67, No. 12, 2409–2424, http://dx.doi.org/10.1080/17470218.2014.931445

Routledge
Taylor & Francis Group

Posture-based processing in visual short-term memory for actions

Staci A. Vicary[1] and Catherine J. Stevens[1,2]

[1]The MARCS Institute, University of Western Sydney, Penrith, NSW, Australia
[2]School of Social Sciences and Psychology, University of Western Sydney, Penrith, NSW, Australia

Visual perception of human action involves both form and motion processing, which may rely on partially dissociable neural networks. If form and motion are dissociable during visual perception, then they may also be dissociable during their retention in visual short-term memory (VSTM). To elicit form-plus-motion and form-only processing of dance-like actions, individual action frames can be presented in the correct or incorrect order. The former appears coherent and should elicit action perception, engaging both form and motion pathways, whereas the latter appears incoherent and should elicit posture perception, engaging form pathways alone. It was hypothesized that, if form and motion are dissociable in VSTM, then recognition of static body posture should be better after viewing incoherent than after viewing coherent actions. However, as VSTM is capacity limited, posture-based encoding of actions may be ineffective with increased number of items or frames. Using a behavioural change detection task, recognition of a single test posture was significantly more likely after studying incoherent than after studying coherent stimuli. However, this effect only occurred for spans of two (but not three) items and for stimuli with five (but not nine) frames. As in perception, posture and motion are dissociable in VSTM.

Keywords: Visual short-term memory; Body movement; Body form; Visual perception; Capacity; Dance.

Movement of the human body is one of the most pervasive visual images observed within the daily environment. Accordingly, body movement perception has been widely researched (see Blake & Shiffrar, 2007, for review). This research has established the features that enable imitation and reproduction of body movements that have been observed visually (e.g., Bird, Osman, Saggerson, & Heyes, 2005; Heyes & Foster, 2002; Stefan et al., 2005). By contrast, the features that enable recognition of movements from visual short-term memory (VSTM) are yet to be described. Body movement is defined as a sequence of successive changes in body posture over time and space (Adshead, 1987). Given this definition, VSTM of observed action might be expected to contain both the complete movement phrase (the "action") and the individual body postures of which the action is composed. As yet, research has not teased apart memory for human action in this way. In two

Correspondence should be addressed to Staci Vicary, The MARCS Institute, University of Western Sydney, Locked Bag 1797, Penrith, NSW 2751, Australia. E-mail: s.vicary@uws.edu.au

This research used stimuli recorded and provided by Beatriz Calvo-Merino. We thank Elizabeth Grima, Kathy Khuu, Karen Lautier, Jessica Menzies, Irakli Moseshvelli, and Jonathon Pye for assistance in data collection. We also thank the Music Cognition and Action research group at the MARCS Institute and two reviewers for comments on an earlier version.

This research was supported by an Australian Postgraduate Award held by Staci Vicary.

experiments, we ask whether body form is remembered after action observation. Drawing on neuroimaging and behavioural research, it was hypothesized that observation of coherent action-like stimuli would lead to poorer rates of recognition of individual action postures from VSTM, suggesting that actions are not stored with direct reference to body form in memory. In supporting this hypothesis, we demonstrate that memory for body movement is not just memory for postures, but rather memory for postures in motion.

Perception of body movement: Form plus motion processing

In a computational model of movement perception, Giese and Poggio (2003) proposed that perception of human action relies on two primary neural pathways, one specialized for processing of body form and another for processing of the body in motion. Neuroimaging studies of movement observation support this model, demonstrating that the perception of body movement involves additional neural structures to those that are involved in the perception of body form. That is, perception of body movement is more than just perception of many changes in body shape. For example, visual observation of both body form and body movement have been specifically associated with regions of medial temporal and fusiform gyri known as extrastriate and fusiform body areas, respectively (EBA, FBA; Downing, Chan, Peelen, Dodds, & Kanwisher, 2006; Downing, Jiang, Shuman, & Kanwisher, 2001; Peelen & Downing, 2007; Pourtois, Peelen, Spinelli, Seeck, & Vuilleumier, 2007; Urgesi, Berlucci, & Aglioti, 2004). However, observation of body movement is also strongly associated with activity in a network of neural regions that are not typically activated by static body form alone. This includes motion sensitive area MT+, as well as a posterior portion of superior temporal sulcus (pSTS) and neurons within inferior parietal lobule and inferior frontal gyrus (see Cross, Mackie, Wolford, & De C Hamilton, 2010, for review). In particular, responses in pSTS appear specific to human movement even when overt body form cues have been removed such as through the use of point light (PL) stimuli (Grossman & Blake, 2001). In PL stimuli, the shape of the body and its movement through space are indicated only by a set of markers placed on major body joints. While individual static frames of PL stimuli appear to display nothing but a random arrangement of dots, presentation of successive PL frames elicits a sense of biological motion (Blake & Shiffrar, 2007). From PL stimuli, observers are reliably able to identify the action being performed despite the lack of overt body cues present (Dittrich, 1993). That is, movement perception is robust even when the shape of the body is absent.

Given that perception of body movement is possible in the absence of form cues, it is unlikely that the neural pathways specialized for form and for motion processing are completely dissociable (Christensen & Calvo-Merino, 2013). However, the notion that some separation exists in neural processing of body form and motion suggests that perception of body movement may not necessarily lead to the retention of body posture. That is, during observation of a body movement, direct processing of individual forms may not occur, especially if they are not necessary in order to discriminate or identify the action. Indeed, even when static images of body form are the only stimuli available to participants, motion remains the primary perception. For example, when viewing a set of static snapshots of successive body shapes, the action that connects each body shape is typically perceived rather than a set of disparate images (Chatterjee, Freyd, & Shiffrar, 1996; Shiffrar & Freyd, 1990, 1993). Despite the lack of actual motion within these stimuli, apparent motion between static images of body form has been shown to influence subsequent dynamic judgements such as perceived trial and item duration (Orgs, Betsmann, Schuur, & Haggard, 2011; Orgs & Haggard, 2011). Likewise, after observing single pictures that imply motion (i.e., an imaging of a body falling to the ground), recognition judgements are biased towards the implied end position (Freyd, 1983; Verfaillie & Daems, 2002). Thus, motion perception occurs in the absence of form cues (as in PL stimuli) and when static images of body form are the only stimuli available.

Therefore, when considering memory for action, if body form is not directly encoded during action observation or is overridden by the perception of motion, then body form may not be retained as part of the primary action representation.

Recognition of body posture after movement observation

The term visual short-term memory (VSTM) refers to the neurocognitive network involved in the temporary storage and manipulation of task-relevant stimuli (Luck & Vogel, 2013; Zimmer, 2008). Recent theories of VSTM suggest that the way in which stimuli are initially perceived will impact upon the way in which they are retained (Luck, 2008). During observation, actions primarily activate motion-processing pathways while form pathways specialized for body shape processing are typically less active. Therefore, if processing type influences retention format, then memory for action may be primarily motion based, with less detailed representation of body form.

Research targeting VSTM has primarily focused on abstract, static objects or objects exhibiting non-biological motion (Papenmeier, Huff, & Schwan, 2012). As a result, relatively little is known about how complex visual stimuli, such as dynamic human actions, are retained. From research that has considered visual memory for action (for a review see Wood, 2007, 2008) there is evidence that action category (i.e., jump, turn), action duration, and action orientation can be retained in VSTM, along with other peripheral features such as actor identity and the spatial location in which the action was performed. Yet, there has been little consideration of whether actions and their constituent body postures are treated similarly in VSTM. Can observers recognize an action they have just seen based solely on static presentations of body shape? This might be the situation, for example, when attempting to identify an observed action by looking only at a static photograph. Previously, we have demonstrated that memory for static postures and dynamic actions differ and may rely on dissociable processing networks (Vicary, Robbins, Calvo-Merino & Stevens, 2014). In a change-detection task, participants recognized actions from static images above the level expected by chance. However, rates of recognition were significantly higher when study and test format matched (i.e., participants observed a set of actions at study and responded to an action at test or observed a set of postures at study and responded to a posture at test). In order to gain a comprehensive understanding of how actions are remembered, it is necessary not only to consider memory at the level of the action, but also to consider the memory for the individual forms, or body shapes, that comprise the action

Does observation of action result in retention of action posture?

If an action posture can be accurately identified as belonging to an observed dynamic action, this would suggest that memory for dance-like movement incorporates both action and form. To investigate memory for posture, the stimuli used in the neuroimaging experiment reported by Downing, Peelen, Wigget, and Tew (2006) were adapted. Using functional magnetic resonance imaging (fMRI), Downing et al. recorded blood-oxygen-level-dependent (BOLD) response within form-based EBA and motion-based pSTS neural regions (amongst others) to "dynamic" stimuli composed entirely of static images of body posture. Two passive viewing conditions were compared by Downing et al.: a coherent condition in which 15 static frames of a single action were presented successively in the correct temporal order, and an incoherent condition in which 15 static frames of different actions were presented successively in a random temporal order. The coherent condition gave the impression of action and elicited a significantly greater BOLD response in the motion-based pSTS than in the form-based EBA region. This is despite the fact that the stimuli were "multi-static" with no explicit motion between successive static images. Alternatively, the incoherent condition elicited a significantly greater response in form-based EBA and FBA than in the motion-based pSTS region, suggesting that these regions were not stimulated by dynamic motion but by the presence of body posture within the stimulus. Downing and colleagues concluded that EBA responds primarily to the presence of human form, while pSTS is sensitive to both the presence of human form and the dynamic relationship between the forms comprising an action.

By adding a recognition phase to the task used by Downing et al. (2006) we can begin to determine, in a behavioural study, whether observation of action leads to retention of action form—that is, whether participants can recognize action posture (form) after processing the coherent, motion-based stimuli. In the current experiments, participants observed sets of coherent and incoherent "multistatic" items and then identified a single test posture as being either "old" (a posture embodied with an action in the study set) or "new" (a posture not embodied within any of the actions in the study set). Observation of coherent multistatic items leads to primarily motion-based "action" processing and, as a result, may lead to limited memory for individual body forms. If so, recognition of action posture should be relatively poor after observation of coherent multistatic items. Alternatively, as the incoherent multistatic items lead to primarily form-based processing, recognition of individual action postures should be relatively good after observation of incoherent multistatic items.

Three important changes were made to the design adapted from Downing et al. (2006) in addition to the inclusion of a recognition phase. First, dance-like actions were used to create the multistatic stimuli. Dance involves arbitrary, intransitive movements of the whole body (Calvo-Merino, Glaser, Grèzes, Passingham, & Haggard, 2005). Unlike grasping and imitative actions, dance-like actions do not require manipulation of external targets or interaction with distinct spatial locations. Therefore, memory for the movement stimuli should be based entirely on action-based cues, rather than cues from external targets. By using dance-like actions, "novice" participants who are unlikely to have any experience in performing the movements can be sampled. In sampling from this novice population, the effects of prior motor experience on action observation of dance actions are limited, pulling the focus of the experiment more tightly onto visual memory alone.

Second, in Downing et al. (2006), coherency was manipulated across items with the incoherent items consisting of frames of several different actions. This meant that while the frames within a single coherent item showed postures that naturally belonged together, the frames within a single incoherent item did not. Therefore, both coherency and context were manipulated. This may have inflated the differences between coherent and incoherent conditions because the coherent items contained implied motion between body shapes that naturally belong together while the incoherent items lacked both implied motion and body shapes that naturally belong together. In the current experiment, coherency was manipulated within items, removing confounds of context. Here, the coherent and incoherent items consist of the same static postures taken from a single action. For the coherent items, the frames appear in the correct temporal sequence, and for the incoherent items the frames appear in an incorrect temporal sequence. This ensures that both conditions contain the same visual information and differ only in the relationship between one frame and the next.

Third, the way in which actions are retained is likely to be influenced by the limited capacity of VSTM. Typically, VSTM capacity for objects is estimated at around four items and even fewer when the target items involve many (Cowan, 2001), compared to few, different features, which must be integrated for retention (Luck, 2008; Wheeler & Triesman, 2002). When the targets are visually presented body movements, a VSTM system dissociable from that used to retain objects is recruited, supporting retention of two to three complete actions both alone and during simultaneous retention of other visual objects (Cortese & Rossi-Arnaud, 2010; Wood, 2007, 2008, 2011). In the current experiments, the influence of capacity on form-based processing of actions is tested in two ways, comparing capacity at both the level of the item and the level of the item components. First, posture recognition is compared for coherent and incoherent items consisting of five and nine static frames. Five is the minimum number of frames that are required to convey beginning, end, and goal (test) posture as well as two preparatory (linking) postures. Nine is much closer to the 15 frames used by Downing et al. (2006) and is expected to well exceed capacity. Second, set sizes of two and three study items are compared. It is expected that the advantage of processing incoherent items, over coherent, will be absent when VSTM capacity is exceeded.

EXPERIMENT 1

Aim, design, and hypotheses

The aim of Experiment 1 was to investigate whether recognition of action postures differs after processing coherent and incoherent stimuli, demonstrating that memory for body movement is more than just memory for posture. A 2 (frame progression: coherent, incoherent) × 2 (frame number: 5 or 9) × 2 (set size: 2 or 3 items) design was implemented with set size realized as a between-subjects variable. Accuracy (d') was the dependent variable.

If dynamic body movements and static body form are processed differently, and this influences retention in VSTM, then differences in posture recognition should be expected after observing coherent and incoherent items that promote motion- and form-based processing, respectively. Specifically, posture recognition should be relatively poor after observing coherent items and relatively good after observing incoherent items. However, this effect is likely to be attenuated when the upper limits of VSTM capacity are reached. It is predicted that the advantage on recognition gained from the processing of incoherent, over coherent, items should be attained in the five-frame and set size 2 conditions, but not in the nine-frame and set size 3 conditions.

Method

Participants

Sixty students enrolled in first year psychology at the University of Western Sydney ($M = 22.85$ years, $SD = 7.4$; 13 male) participated in return for course credit. Participants were dance-novices (no reported professional dance training) and reported normal or corrected-to-normal vision. All participants were naive to the task.

Materials and equipment

Dance-like action stimuli consisted of a set of 10 ballet items performed by an experienced male ballet dancer, used previously in Vicary et al. (2014) and Calvo-Merino et al. (2005). The items were each 3 s in length, and each consisted of one performer executing a single ballet item. All movements began from roughly the right-hand side of the screen (stage left for the performer) and travelled across the screen to the left. In this instance, memory for the most pertinent posture was investigated. Specifically, we investigated recognition of the "goal posture", which was defined as the body shape representing both the maximal deviation from the neutral start position and the height or peak of the action phrase. For example, this might include the full extension of the legs and the arms achieved in a grand jeté. Images of the test postures are shown in Figure 1.

Criteria were established to select the frames that would be used in the five- and nine-frame stimuli. Specifically, the first and last frames were always selected from the complete item, as was the frame representing the goal posture. In the nine-frame condition the test-frame always represented Frame 5, 6, or 7, whilst in the five-frame condition, the test-frame was always Frame 3 or 4 (this varied across the items). The remaining six frames were classed as "preparatory frames", representing preparative or transitional motions connecting the beginning, peak, and final postures. For example, the five-frame condition consisted of the beginning, goal (test), and final frames plus two preparatory frames, while the nine-frame condition consisted of the beginning, goal (test), and final frames plus six preparatory frames.

Coherent items consisted of the frames (five or nine) progressing in the correct sequential order, from the beginning to the end of the action. Individual frames were played sequentially with no delay or blackout between one frame and the next. The incoherent items consisted of the frames (five or nine) progressing in the incorrect sequential order, violating the natural progression from the

Figure 1. *Complete set of test stimulus items. To view this figure in colour, please visit the online version of this Journal.*

beginning to the end of the action (see Figure 2 for an example of coherent and incoherent items). Criteria were established for creating the incoherent stimuli. Specifically: (a) two frames could not play in natural succession (e.g., Frame 3 could not follow Frame 2); (b) a frame could not appear in the same place as it naturally would in the coherent condition (e.g., Frame 3 could not appear in Position 3); and (c) frames showing very similar postures could not occur successively (this was a qualitative decision made by the first author to ensure that the incoherent items did not appear to flow naturally at any time). The one exception to these criteria was that for all items the test-frame would appear in its natural position; that is, if Item 1 has the test-frame at Position 3 in the coherent condition, then in the incoherent condition the test-frame also appeared in Position 3.

Figure 2. *An example of a single multistatic action used in the study phase. A and B are examples of coherent condition items; A shows a five-frame item, and B shows a nine-frame item. C and D are examples of incoherent condition items; C shows a five-frame item, and D shows a nine-frame item. To view this figure in colour, please visit the online version of this Journal.*

Five- and nine-frame stimuli were also matched in the incoherent stimuli such that the "incoherency" of the frames was similar across the two frame numbers. Each stimulus item had a 3000-ms duration; therefore the study phase lasted 9000 ms. The stimuli shown in the test phase of the experiment consisted of a single frame of the test-posture. The test-postures were presented for 3000 ms at test, after which a blank screen was displayed.

Stimuli were edited from original high-definition files in Adobe Premiere Suite and were played back during the experiment as .avi files. The experiment was programmed and run on LG PCs via DMDX software, with a refresh rate of 16.4 and video resolution of 1280 × 768.

Procedure

Participants signed informed consent in line with the University of Western Sydney's Human Research Ethics Committee Guidelines. Participants were instructed that they would observe sets of two or three dance-like actions that might appear "jerky" and sometimes out of order, but that they were to take note of what action was being performed regardless of its form.

Each trial progressed in the following way: A blank screen showing the word "study" signalled the beginning of a trial. Subsequently, two or three dance-like actions appeared sequentially in the centre of the screen. On any given trial, all study items were either coherent or incoherent and contained either five or nine frames. That is, three incoherent items or three coherent items were shown at study, never a combination of the two stimulus formats. The word "test" was then displayed in the centre of the screen, and a single test-frame appeared. The test-frame could be either old (representing an action from the study set) or new (not representing an action from the study set). Participants then responded to the test-frame as "old" or "new" using the computer's shift keys. Responses were time constrained during the experiment, with a total response period of 5 s, including the 3-s duration of the test item. After a response was recorded, a new trial began (see Figure 3 for an example of a trial). Eight practice trials with corrective feedback were completed before the task began. The same ballet movements as those used in the main experiment were used as the practice items. This allowed participants to familiarize themselves with the type of items they would expect in the task and also provided a basis on which to identify participants who had misunderstood or were unable to perform the task. In total, 40 experimental trials were completed, allowing 10 trials for each combination of frame number and frame progression. Within each set of 10 experimental trials, five trials contained a "new" action key-frame at test, while five trials contained an "old" action key-frame at test. The experiment took approximately 35 min to complete.

Results and discussion

Data analysis was conducted using d' as an overall measure of accuracy. Descriptive statistics, including individual hit and false-alarm rates, are shown in Figure 4 and Table 1. A 2 (frame progression; coherent, incoherent) by 2 (frame number; five, nine) by 2 (set size; two, three items) mixed measures analysis of variance (ANOVA) was performed on d' values, with planned comparisons to test specific hypotheses.

With alpha set at .05, there was a significant main effect of set size, $F(1, 58) = 3.08$, $p = .019$, $\eta_p^2 = .09$, a significant two-way interaction between frame progression and set size, $F(1, 58) = 13.04$, $p = .001$, $\eta_p^2 = .18$, and significant three-way interaction amongst all variables, $F(1, 58) = 1.81$, $p = .014$, $\eta_p^2 = .10$. The three-way interaction demonstrated differences in accuracy depending on set size. Thus, two repeated measures ANOVAs investigated effects of frame progression and frame number separately for the set size factor.

For the Set Size 2 condition, the main effect of frame progression was significant, $F(1, 29) = 6.63$, $p = .015$, $\eta_p^2 = .19$, indicating greater overall accuracy in the incoherent condition ($M = 1.38$, $SD = 0.67$) than in the coherent condition ($M = 1.05$, $SD = 0.61$). The main effect of frame number was not significant, $F(1, 29) = 1.07$, $p = .11$, $\eta_p^2 = .09$, with accuracy in the five-frame condition ($M = 1.3$, $SD = 0.60$) similar to that in the nine-frame condition ($M = 1.13$, $SD = 0.68$). However, the

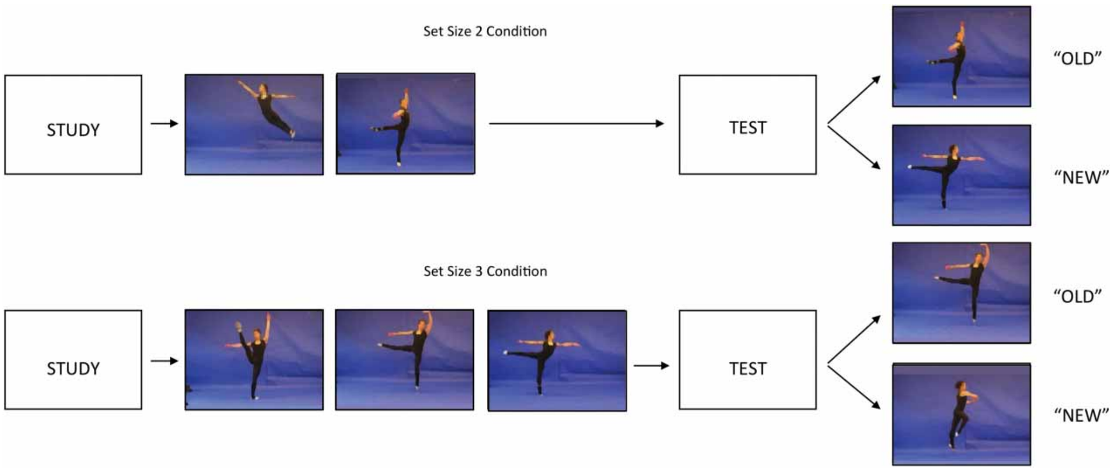

Figure 3. *Trial progression shown separately for Set Size 2 and Set Size 3 conditions. In the study phase, participants observed two or three multistatic actions, depending on which condition they had been assigned to. Each multistatic action had a duration of 3 s and was, on differing trials, composed of either five or nine frames in coherent or incoherent order (see Figure 2). Subsequently, a single static test posture was shown. This test posture was either old (included in the study phase) or new (not included in the study phase). Test postures had a duration of 3 s. The total duration of the response period (including presentation of the test posture) was 5 s, after which the trial was terminated, and a new trial began. To view this figure in colour, please visit the online version of this Journal.*

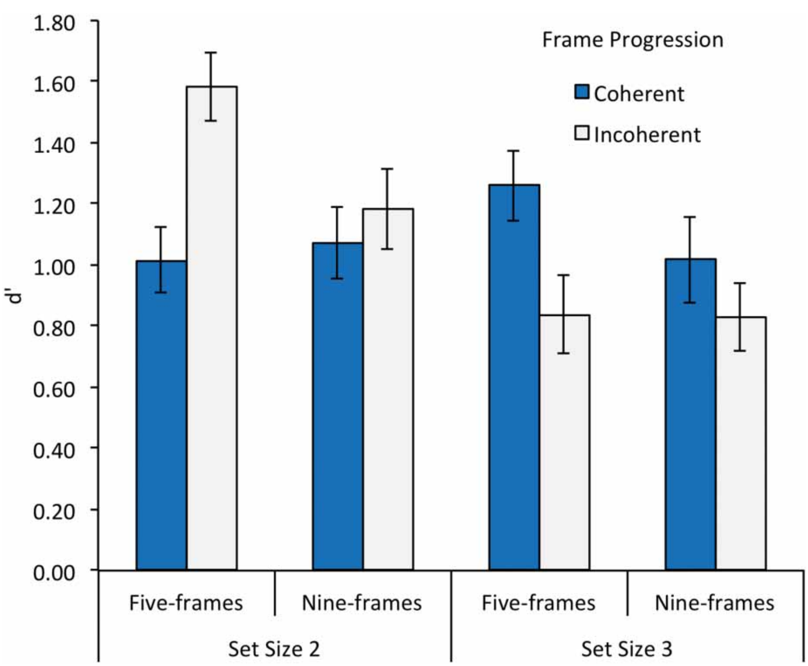

Figure 4. *Experiment 1: Mean* d′ *for posture recognition for frame progression, frame number, and set size conditions. Bars represent standard error of the mean. To view this figure in colour, please visit the online version of this Journal.*

Table 1. *Descriptive statistics*

			Set Size 2			Set Size 3		
Experiment	*Frame number*	*Frame progression*	*Hit rate*	*False-alarm rate*	d′	*Hit rate*	*False-alarm rate*	d′
Experiment 1	Five frames	Coherent	.69 (.19)	.31 (.18)	1.02 (0.58)	.68 (.17)	.21 (.15)	1.26 (0.63)
		Incoherent	.71 (.19)	.11 (.15)	1.58 (0.62)	.60 (.21)	.29 (.19)	0.84 (0.70)
	Nine frames	Coherent	.68 (.19)	.28 (.15)	1.07 (0.64)	.63 (.21)	.25 (.18)	1.02 (0.77)
		Incoherent	.55 (.25)	.10 (.12)	1.18 (0.71)	.51 (.17)	.19 (.15)	0.83 (0.61)
Experiment 2	Five frames	Coherent	.65 (.25)	.26 (.18)	1.03 (0.63)	.60 (.16)	.24 (.19)	0.98 (0.65)
		Incoherent	.63 (.19)	.14 (.16)	1.30 (0.56)	.60 (.26)	.25 (.20)	0.94 (0.65)
	Nine frames	Coherent	.60 (.21)	.24 (.15)	0.98 (0.65)	.65 (.21)	.29 (.21)	0.98 (0.64)
		Incoherent	.59 (.25)	.20 (.20)	1.04 (0.62)	.58 (.22)	.25 (.19)	0.89 (0.58)

Note: The numbers in brackets are the Standard Deviations.

effect of frame progression was not the same for both levels of frame number, $F(1, 29) = 9.42$, $p = .005$, $\eta_p^2 = .25$. It was hypothesized that when the stimuli are composed of five frames, posture recognition will be more accurate in the incoherent than in the coherent condition, although when the stimuli are composed of nine frames, accuracy in the incoherent and coherent conditions may be similar. Consistent with this hypothesis, when the stimuli consisted of five frames, accuracy in the incoherent condition ($M = 1.58$, $SD = 0.62$) was significantly greater than accuracy in the coherent condition ($M = 1.02$, $SD = 0.58$), $t(29) = 4.14$, $p < .001$, $r = .42$. Further, as predicted, when the stimuli consisted of nine frames, accuracy in the incoherent ($M = 1.18$, $SD = 0.71$) and coherent ($M = 1.07$, $SD = 0.64$) conditions did not differ, $t(29) = 0.667$, $p = .51$, $r = .08$.

For the Set Size 3 condition, a significant main effect of frame progression was also obtained, $F(1, 29) = 6.43$, $p = .017$, $\eta_p^2 = .19$. Accuracy for the coherent condition ($M = 1.14$, $SD = 0.70$) was significantly greater than accuracy in the incoherent condition ($M = 0.84$, $SD = 0.66$). Contrary to the results of the Set Size 2 condition, there was no interaction between frame progression and frame number, $F(1, 29) = 1.06$, $p = .312$, $\eta_p^2 = .04$. Given capacity limitations of VSTM for actions, it was hypothesized that posture recognition would be similar for the incoherent and coherent stimuli comprising five frames. However, when the stimuli consisted of nine frames, posture recognition after incoherent stimuli should be particularly poor. Contrary to this prediction, planned comparisons indicate that the difference between incoherent and coherent stimuli was significant for the five-frame stimuli but not for the nine-frame stimuli. Specifically, when the stimuli contained five frames, accuracy in the coherent condition ($M = 1.26$, $SD = 0.63$) was significantly greater than accuracy in the incoherent condition ($M = 0.84$, $SD = 0.70$), $t(29) = 2.83$, $p = .008$, $r = .30$. When the stimuli contained nine frames, accuracy in the coherent ($M = 1.02$, $SD = 0.77$) and incoherent conditions ($M = 0.83$, $SD = 0.61$) did not differ, $t(29) = 1.03$, $p = .31$, $r = .14$.

Experiment 1 demonstrates that posture recognition is more accurate after observation of incoherent items than after observation of coherent items. This result supports the hypothesis that observation of action does not necessarily lead to the retention of action posture in VSTM. When action stimuli were manipulated so that plausible biological motion was not apparent (making the action incoherent), participants were able to retain and recognize individual action postures. When these same action postures were observed within the correct temporal sequence (coherent condition), participants were significantly less accurate. However, while this incoherent advantage is evident for multistatic items comprising five individual postures, the advantage is absent both for items comprising nine individual postures and when the size of the study set is increased from two to three items. This lack of an incoherence advantage may be attributed to overall poorer posture-based memory in these conditions, as evidenced by the significant decrease in accuracy for set sizes of three, compared to two items, and generally poorer accuracy for items with nine than for those with five frames overall (effect approaches significance at $p = .1$). Thus, it appears that processing actions in a posture-based manner is only an advantage when the number of items or item parts falls within or close to VSTM capacity. However, with regard to the nine-frame stimuli, another explanation is also possible. In order to control stimuli duration, in Experiment 1 both the five- and nine-frame stimuli were controlled to an overall duration of 3 s. This meant that each individual frame was viewed for less time in the nine-frame condition than in the five-frame condition. Therefore, the difference between recognition rates for the five- and nine-frame stimuli may be due to reduced exposure to each individual frame of the nine-frame relative to the five-frame stimuli. While differences in the duration of actions may not be key to differences in retention patterns in VSTM (Wood, 2007), it was important to nonetheless rule out frame-rate differences as an explanation for the difference between five- and nine-frame conditions. Therefore, Experiment 2 was conducted as a replication of Experiment 1, with frame duration (rather than item duration) matched across conditions. If the lack of difference between coherent and incoherent stimuli in the nine-frame condition is due to capacity, not unequal frame duration, results should replicate those of Experiment 1.

EXPERIMENT 2

Method

Participants

Forty-four students enrolled in first-year psychology at the University of Western Sydney ($M =$ 24.26 years, $SD = 1.03$ years; 6 male) volunteered to participate in return for course credit. All participants were dance novices who were naive to the task and had not participated in Experiment 1.

Materials and equipment

Materials and equipment were as those in Experiment 1, with the only change being to the duration of the nine-frame stimuli items. In Experiment 2, the duration of each frame, across both five- and nine-frame conditions, was 600 ms. Thus, the five-frame stimuli had a total duration of 3 s, and the nine-frame stimuli had a total duration of 5.4 s.

Procedure

Procedure was the same as that in Experiment 1.

Results and discussion

An omnibus $2 \times 2 \times 2$ mixed ANOVA produced no significant main effects or interactions. However, consistent with Experiment 1, in the Set Size 2 condition, accuracy was significantly greater in the incoherent five-frame condition ($M = 1.30$, $SD = 0.56$) than in the coherent five-frame condition ($M = 1.03$, $SD = 0.63$), $t(21) = 1.71$, $p = .046$, $d = 0.53$. By contrast there was no significant difference between accuracy for the incoherent nine-frame stimuli ($M = 1.04$, $SD = 0.62$) and that for the coherent nine-frame stimuli ($M = 0.98$, $SD = 0.65$), $p = .67$ (see Figure 5 and Table 1).

For Set Size 3, the difference between accuracy in the five-frame condition for coherent ($M =$ 0.98, $SD = 0.65$) and incoherent stimuli ($M =$ 0.94, $SD = 0.65$) was not significant, $p = .42$. Similarly, accuracy in the nine-frame condition did not differ significantly for the coherent ($M =$ 0.98, $SD = 0.64$) and incoherent ($M = 0.89$, $SD = 0.58$) stimuli, $p = .69$.

Replicating results from Experiment 1, Experiment 2 demonstrates better rates of posture recognition after observing two incoherent than after observing two coherent items. This effect is only observed when the stimuli are made up of five, but not nine, static frames. Importantly, in Experiment 2, the frame duration of five- and

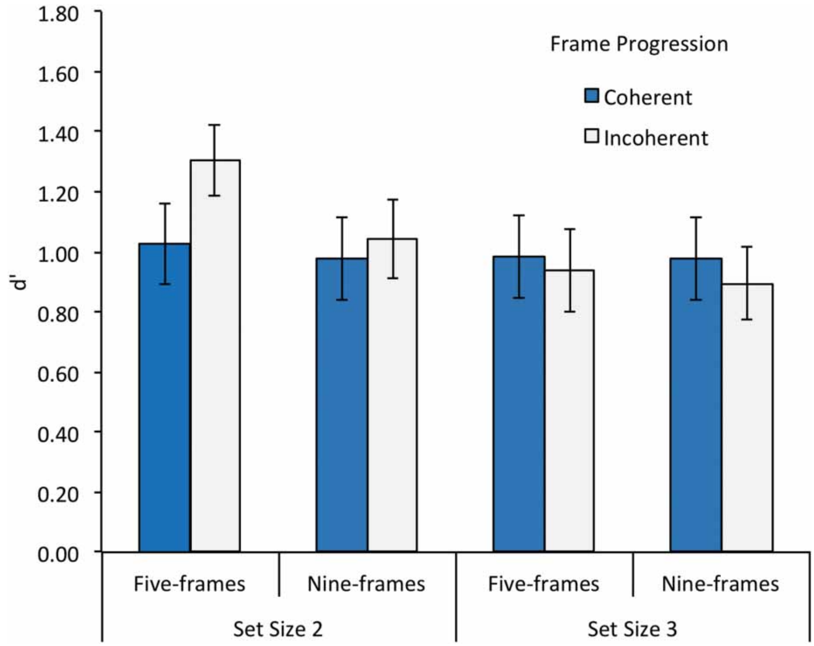

Figure 5. *Experiment 2: Mean* d′ *for posture recognition for frame progression, frame number, and set size conditions. Bars represent standard error of the mean. To view this figure in colour, please visit the online version of this Journal.*

nine-frame stimuli was matched such that each frame was observed for an equal amount of time regardless of how many frames were observed overall. The lack of an advantage for posture recognition from incoherent items consisting of nine frames should therefore not be attributed to differences in frame duration. It is more likely that, with nine frames per item, VSTM capacity has been exceeded, leading to poorer posture recognition.

GENERAL DISCUSSION

Two experiments were designed to investigate whether VSTM for action necessarily entails retention of action posture. More specifically, the accuracy with which body postures can be recognized after observation of multistatic items that do or do not resemble natural action was measured. The format of action stimuli was manipulated following Downing et al. (2006) such that the postures required to execute a single action were observed in the correct (coherent) or incorrect (incoherent) temporal order. Coherent items were presented to elicit primarily motion-based processing, while incoherent items were presented to elicit primarily posture-based processing. It was hypothesized that posture recognition should be facilitated by the incoherent items because they promote perception of individual postures. Conversely, posture recognition should be inhibited by the coherent items as they promote perception of the action in a more coherent or holistic manner. In both experiments, posture recognition was significantly greater in the incoherent than in the coherent trials, supporting the hypothesis that processing of actions does not entail retention of action posture. However, this effect was only upheld for items with five (but not nine) frames, and when study sets comprised a total of two (but not three) total items. Thus, the bounded capacity of VSTM limits the degree to which action form can be maintained after observing action-like items.

This is the first experiment that has explicitly investigated memory for body posture after perception of action stimuli. The findings support research on memory for action, demonstrating that memory is relatively good for a set size of around two items (Cortese & Rossi-Arnaud, 2010; Wood, 2007, 2008). The results are novel in demonstrating that posture or body shape is retained after observation of action stimuli and to varying degrees of accuracy depending on the type of processing that has taken place.

The present results demonstrate that observation of action does not necessarily entail retention of action posture. The type of processing taking place during observation of visual stimuli has previously been demonstrated to influence the form in which the stimuli are maintained in visual memory (Engle, 2010; Slotnick & Thakral, 2011). In the present experiments, processing of the incoherent items led to greater rates of posture recognition than did processing of the coherent items. This is despite the fact that both stimulus types consisted of the same overall static body images and differed only in the order in which these images were presented. Thus, the processing of postures out of order appears to have enhanced posture retention in VSTM, relative to the processing of postures in order.

A likely reason for this difference between recognition of postures in the coherent and incoherent conditions is the tendency to observe and retain motion in the coherent condition. This is to say that processing postures in motion led to retention not of postures but of action. Prior research on apparent motion demonstrates that biomechanically plausible motion can be perceived from completely static images portraying an action, as long as the temporal properties of the action are maintained (i.e., the interstimulus interval is not too short; Chatterjee et al, 1996; Shiffrar & Freyd, 1990). The stimuli used in the present experiments are quite different to those typically shown to elicit apparent motion. Nonetheless, motion appears to have been perceived in the coherent condition such that coherent items do not seem to have been processed as a multistatic set of postures, but rather as a single biological action. This is demonstrated not only in the higher rates of recognition coming from the incoherent than from the coherent condition but also in the responses given by participants to a posttest questionnaire. In these questionnaires, participants were asked to describe how they attempted to maintain each item within the task. For the coherent items, responses often included such statements as "I tried to remember the movement as a whole", "I tried to remember what actions they were doing", and "(I tried to remember) the sequence of movements and their flow". Alternatively, for the incoherent items responses often included such statements as "(remembering) what body parts were used", "I tried to remember the different shapes made with the body, e.g., tall and thin", and "Looking at the main movement that the others lead to and then reordering them". Together with the accuracy measure, these statements support the conclusion that coherent items were not processed as postures, but as motion.

The notion that postures and postures in motion may differ in VSTM is informative when extending theories of action perception into theories of action memory. The experiments reported here demonstrate that, as in perception, memory for body form and memory for body movement differ. When attempting to retain two actions with the goal of later recognizing a single static body posture, participants were most accurate if apparent motion was not induced. That is, recognizing form was easier when motion had not initially been implied, allowing form-based processing to take place. Alternatively, when a sense of implied motion was present during study, form-based processing seems to have been reduced in favour of motion-based processing. Thus, as suggested by Orgs, Kirsch, and Haggard (2013), motion processing appears to be preferenced over form processing, even when viewing purely static images. Orgs and colleagues suggest that this might occur via a biological motion recognition system wherein the network of brain regions sensitive to successive body postures activates a dynamic percept of action in place of static postural information. While form is not lost or replaced, it is often substituted in favour of a more useful percept of motion (see Orgs et al., 2013). The experiments presented here demonstrated that the perception of motion over form carries through to working memory with implications for the type of recognition judgements that can later be made.

In the context of prior research, the results of the current experiments suggest that the process by which actions are retained in memory is one that enables separable storage of the complete action in addition to action form. Given that, in daily life, memory for action is more likely to be necessary than memory for action postures, motion processing probably overrides less crucial posture-based processing. When processing action-like stimuli, such as the coherent items in Experiments 1 and 2, attention is paid to the overall motion, with less emphasis on the individual shapes made with the body. Only when the motion is made less prominent, such as with the incoherent items, does processing of body form take priority. This preferential processing of motion is demonstrated both in neuroimaging research reporting higher BOLD responses in motion regions than in form regions during action observation (Downing et al., 2006) and behavioural research demonstrating relatively poor rates of posture recognition after action observation (Vicary et al., 2014). The new finding here is the demonstration of an apparent inability to override motion processing in favour of posture-based processing in some conditions, even when the latter is directly task relevant. As a result, posture recognition is more likely when the action retention network has been minimally activated during perception.

Finally, in Experiment 1 a significant interaction was obtained with accuracy greatest after observation of the coherent multistatic stimuli in the Set Size 3 condition. This result is contrary to hypotheses but is still consistent with an explanation of posture-based processing in a capacity-limited visual memory system. This is because the primary change in the pattern of accuracy between Set Sizes of 2 and 3 in Experiment 1 is not necessarily improved performance for coherent trials, but rather much poorer performance on the incoherent trials at the larger set size. This difference is driven by an increase in the number of false alarms (new items mistaken for old) made on the incoherent trials as the number of items in the study set was increased (see Table 1). It has been argued here that participants engage in posture-based processing when viewing the incoherent items. While this posture-based processing appears to have assisted recognition of postures from a set of two items, it appears to have impaired recognition of postures from a set of three items. With three items at study, the number of individual postures processed is likely to be too large for the limited-capacity VSTM system to retain. Therefore, even though posture-based processing may have occurred at study, it does not assist recognition when there are many items to remember.

The reversal of the coherency effect observed in Experiment 1 was not replicated in Experiment 2 (accuracy did not differ between coherent and incoherent conditions for set size of 3) but marks an interesting point for future research. It is possible that for the coherent items, participants were able to rely on gist-based representations of the action goal (for example, remembering that the actions were a “jump”, “turn”, and “kick”) and work backwards at test to determine whether or not the test posture “fits” within these categories. With increased exposure to each action posture over the course of Experiment 2 (due to the increased duration of the nine-frame items), the generalization and retention of action gist may have been less effective than it was in Experiment 1. As a result, participants make slightly more errors (fewer hits and more false alarms) on coherent trials at set size 3 in Experiment 2 (see Table 1), and the coherency reversal observed in Experiment 1 does not recur. This is speculative, but the notion of gist-based representations has in part been addressed by Urgolites and Wood (2013) in experiments aiming to determine the fidelity of action representations in visual memory. In a series of experiments, Urgolites and Wood demonstrate that participants are able to recognize observed actions when presented within organized categories (i.e., recognizing a jump presented in the context of other jumps) and identify the correct range of motion that was performed. This suggests that actions are retained with fine-grained detail and may not be simple gist representations. However, the question remains as to the accuracy of memory for observed body postures within an action that involves many changes in form across time and space, such as dance. Research currently underway considers the accuracy of posture recognition both across and within the multistatic type actions utilized here. For posture recognition *within* an action, reliance on "gist" type representations is rendered impractical. Thus, ongoing research will build on the results presented here to further expand our understanding of how actions and action postures are retained in VSTM.

Implications

The results of the two experiments reported here also have important ecological implications for theories of motor learning and social interaction. Given the action-based demands of typical interpersonal interaction, the notion that retention of action is preferential over action posture is parsimonious. However, in domains as vast as motor learning and witness testimony, the notion that "the action as a whole is more important than the configurations of which the action is composed" may not be entirely correct. For example, in learning dance, the goal is often for the learner to retain visual memory of individual postures that the teacher has embodied. In this instance, given the results of the present experiment, efforts should be made to demonstrate the action postures both in the presence and in the absence of the overall motion within which they will be embedded. Additionally, if the observer needs to remember many actions, this will come at the cost of being less able to remember, or, more specifically, recognize, individual body postures or shapes of each action within the set. The set size effect of capacity indicates that when the upper limits of capacity have been reached, memory for individual action posture will be limited. Therefore, observational learning of action and memory for discrete action posture is best done in blocked phases of small action sets (around two) to best facilitate memory for postures and postures in motion.

Research to date has demonstrated that observation of bodies and body movement activates a neural network specialized for detection of biological form and/or biological motion. The results of the current experiments demonstrate that this network extends to VSTM with action and action posture also retained via dissociable mechanisms. Although body movement comprises many postures performed over space and time, perception and retention of action do not necessarily mean perception and retention of action posture.

Supplemental material

The Supplemental material contains examples of the different conditions and is available via the "Supplemental" tab on the article's online page (http://dx.doi.org/10.1080/13506285.2014.931445).

Original manuscript received 4 October 2013
Accepted revision received 23 March 2014
First published online 10 July 2014

REFERENCES

Adshead, J. (1987). *Dance analysis: Theory and practice.* London: Dance Books.

Bird, G., Osman, M., Saggerson, A., & Heyes, C. (2005). Sequence learning by action, observation and action observation. *British Journal of Psychology*, *96*, 371–388. doi:10.1348/000712605X47440

Blake, R., & Shiffrar, M. (2007). Perception of human motion. *Annual Review of Psychology*, *58*, 47–73. doi:10.1146/annurev.psych.57.102904.190152

Calvo-Merino, B., Glaser, D. E., Grèzes, J., Passingham, R. E., & Haggard, P. (2005). Action observation and acquired motor skills: An fMRI study with expert dancers. *Cerebral Cortex*, *15*, 1243–1249. doi:10.1093/cercor/bhi007

Chatterjee, S. H., Freyd, J. J., & Shiffrar, M. (1996). Configural processing in the perception of apparent biological motion. *Journal of Experimental Psychology: Human Perception and Performance*, *22*, 916–929.

Christensen, J. F., & Calvo-Merino, B. (2013). Dance as a subject for empirical aesthetics. *Psychology of Aesthetics, Creativity, and the Arts*, *7*, 76–88.

Cortese, A., & Rossi-Arnaud, C. (2010). Working memory for ballet moves and spatial locations in professional ballet dancers. *Applied Cognitive Psychology*, *24*, 266–286. doi:10.1002/acp

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*, 87–114.

Cross, E. S., Mackie, E. C., Wolford, G., & De C Hamilton, A. F. (2010). Contorted and ordinary body postures in the human brain. *Experimental Brain Research*, *204*, 397–407. doi:10.1007/s00221–009–2093-x

Dittrich, W. H. (1993). Action categories and the perception of biological motion. *Perception*, *22*, 15–15.

Downing, P. E., Chan, a. W. Y., Peelen, M. V., Dodds, C. M., & Kanwisher, N. (2006). Domain Specificity in Visual Cortex. *Cerebral Cortex*, *16*, 1453–1461. doi:10.1093/cercor/bhj086

Downing, P. E., Jiang, Y., Shuman, M., & Kanwisher, N. (2001). A cortical area selective for visual processing of the human body. *Science*, *293*, 2470–2473.

Downing, P. E, Peelen, M. V., Wiggett, A. J., & Tew, B. D. (2006). The role of the extrastriate body area in action perception. *Social Neuroscience*, *1*, 52–62. doi:10.1080/17470910600668854

Engle, R. W. (2010). Role of working-memory capacity in cognitive control. *Current Anthropology*, *51*, S17–S26.

Freyd, J. J. (1983). Representing the dynamics of a static form. *Memory & Cognition*, *11*, 342–346.

Giese, M. A., & Poggio, T. (2003). Neural mechanisms for the recognition of biological movements. *Nature Reviews Neuroscience*, *4*, 179–192.

Grossman, E. D., & Blake, R. (2001). Brain activity evoked by inverted and imagined biological motion. *Vision Research*, *41*, 1475–1482.

Heyes, C. M., & Foster, C. L. (2002). Motor learning by observation : Evidence from a serial reaction time task. *The Quarterly Journal of Experimental Psychology*, *55*, 593–607. doi:10.1080/02724980143000389

Luck, S. J. (2008). Visual short-term memory. In S. J. Luck & A. Hollingworth (Eds.), *Visual memory* (pp. 43–87). Cary, NC: Oxford University Press.

Luck, S. J., & Vogel, E. K. (2013). Visual working memory capacity: From psychophysics and neurobiology to individual differences. *Trends in cognitive sciences*, *17*, 391–400.

Orgs, G., Bestmann, S., Schuur, F., & Haggard, P. (2011). From body form to biological motion the apparent velocity of human movement biases subjective time. *Psychological science*, *22*, 712–717.

Orgs, G., & Haggard, P. (2011). Temporal binding during apparent movement of the human body. *Visual Cognition*, *19*, 833–845.

Orgs, G., Kirsch, L., & Haggard, P. (2013). Time perception during apparent biological motion reflects subjective speed of movement, not objective rate of visual simulation. *Experimental Brain Research*, *227*, 223–229.

Papenmeier, F., Huff, M., & Schwan, S. (2012). Representation of dynamic spatial configurations in visual short-term memory. *Attention, Perception, & Psychophysics*, *74*, 397–415.

Peelen, M. V., & Downing, P. E. (2007). The neural basis of visual body perception. *Nature Reviews Neuroscience*, *8*, 636–48. doi:10.1038/nrn2195

Pourtois, G., Peelen, M. V., Spinelli, L., Seeck, M., & Vuilleumier, P. (2007). Direct intracranial recording of body-selective responses in human extrastriate visual cortex. *Neuropsychologia*, *45*, 2621–2625. doi:10.1016/j.neuropsychologia.2007.04.005

Shiffrar, M., & Freyd, J. J. (1990). Apparent motion of the human body. *Psychological Science*, *1*, 257–264.

Shiffrar, M., & Freyd, J. J. (1993). Timing and apparent motion path choice with human body photographs. *Psychological Science*, *4*, 379–384.

Slotnick, S. D., & Thakral, P. P. (2011). Memory for motion and spatial location is mediated by contralateral and ipsilateral motion processing cortex. *NeuroImage*, *55*, 794–800.

Stefan, K., Cohen, L. G., Duque, J., Mazzocchio, R., Celnik, P., Sawaki, L., … Classen, J. (2005). Formation of a motor memory by action observation. *The Journal of Neuroscience*, *25*, 9339–9346. doi:10.1523/JNEUROSCI.2282–05.2005

Urgesi, C., Berlucchi, G., & Aglioti, S. M. (2004). Magnetic stimulation of extrastriate body area impairs visual processing of nonfacial body parts. *Current Biology*, *14*, 2130–2134. doi:10.1016/j.cub.2004.11.031

Urgolites, Z. J., & Wood, J. N. (2013). Visual long-term memory stores high-fidelity representations of observed actions. *Psychological Science*, *24*, 403–411.

Verfaillie, K., & Daems, A. (2002). Representing and anticipating human actions in vision. *Visual Cognition*, *9*, 217–232.

Vicary, S., Robbins, R., Calvo-Merino, B., & Stevens, C. J. (2014). Recognition of dance-like actions: Memory for static postures or dynamic movements?. *Memory & Cognition*, *42*(5), 755–767.

Wheeler, M. E., & Triesman, A. M. (2002). Binding in short-term visual memory. *Journal of Experimental Psychology: General*, *131*, 48–64.

Wood, J. N. (2007). Visual working memory for observed actions. *Journal of Experimental Psychology: General*, *136*, 639–52. doi:10.1037/0096–3445.136.4.639

Wood, J. N. (2008). Visual memory for agents and their actions. *Cognition*, *108*, 522–32. doi:10.1016/j.cognition.2008.02.012

Wood, J. N. (2011). When do spatial and visual working memory interact?. *Attention Perception and Psychophysics*, *73*, 420–439.

Zimmer, H. D. (2008). Visual and spatial working memory: from boxes to networks. *Neuroscience & Biobehavioral Reviews*, *32*, 1373–1395.