



Brief article

Visual memory for agents and their actions

Justin N. Wood

Harvard University, Department of Psychology, 33 Kirkland Street, Cambridge, MA 02138, USA

Received 11 July 2007; revised 11 February 2008; accepted 26 February 2008

Abstract

Humans spend a considerable amount of time remembering other individuals' actions. Nevertheless, it is unclear how the visual system stores information about the identities of agents and their actions. To address this, I used a change detection method where observers were asked to remember agents and the actions they performed. Results show that observers can maintain information about both 2–3 agents and 2–3 actions simultaneously. However, they are highly impaired for remembering which agent performed which action, indicating that agent and action information are retained separately in visual working memory. Further experiments show that agent and action information can be bound together when the visual input contains the appropriate cues. However, this binding process significantly reduces the total amount of information that can be retained. Together, these results show that (1) an additional, resource-demanding process is needed to integrate agent and action information stored in separate working memory stores, and (2) the extent to which these two types of information are bound into integrated units depends largely on the presence of specific cues in the visual input.

© 2008 Elsevier B.V. All rights reserved.

Keywords: Visual working memory; Visual short-term memory; Actions; Agents

1. Introduction

Humans depend on their ability to remember other individuals and the actions of those individuals. In order to guide behavior after an action has been observed, infor-

E-mail address: jwood@wjh.harvard.edu

0010-0277/\$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.cognition.2008.02.012

Please cite this article in press as: Wood, J. N., Visual memory for agents and their actions, *Cognition* (2008), doi:10.1016/j.cognition.2008.02.012

mation about the action needs to be stored in some type of temporary buffer, known as visual working memory (VWM). To date, however, little is known about how working memory maintains visual information about the identities of agents and their actions.

Neuroimaging studies show that early in visual processing, information about agent identities and actions are supported by distinct neural substrates. For instance, viewing images of the human body activates the lateral occipitotemporal area (Downing, Jiang, Shuman, & Kanwisher, 2001; Peelen, Wiggett, & Downing, 2006), whereas viewing bodily movements activates mirror neuron areas in the prefrontal cortex, an area that is insensitive to the identity of the acting agent (Ruby & Decety, 2001). Moreover, a transcranial magnetic stimulation study provides causative evidence for this dissociation, showing that interference with the lateral occipitotemporal area impairs the discrimination of bodily identity whereas interference with the prefrontal cortex impairs the discrimination of bodily actions (Urgesi, Candidi, Ionta, & Aglioti, 2006). This dissociation between the neural substrates that process information about agent identities and actions raises three central questions regarding how these types of information are maintained by the visual system: (i) Is agent identity information stored in the same working memory system as action information, or are they stored in separate, specialized working memory systems? If identity and action information are stored separately, then (ii) how does the visual system bind identity and action information into integrated representations? (iii) To what extent is this binding process under conscious control?

To address these questions, I used a variant of the change detection procedure (Phillips, 1974) used previously to measure the storage capacity of VWM for observed actions (Wood, 2007) and objects (e.g., Luck & Vogel, 1997; Wheeler & Treisman, 2002). In a previous study, Wood (2007) showed that it is possible to retain between 2 and 3 observed bodily actions in VWM at one time independent of the number of action properties that need to be retained for each action. This indicates that VWM can retain observed actions as integrated representations. Wood (2007) also found that when observers were asked to remember a sequence of actions and then an array of simultaneously-presented colored objects, the action and object information did not compete with each other for working memory resources. This indicates that action and object information are stored separately when they belong to different visual objects. Critically, does the visual system continue to store action and object information separately when the object information defines the identity of the acting agent? Evidence from fMRI suggests that agent and action information are integrated by attentional processes when they belong to the same visual object (O'Craven, Downing, & Kanwisher, 1999), which raises the possibility that it is these integrated agent-action percepts that are stored in VWM. Experiment 1 tests this possibility.

2. Experiment 1

On each trial, observers viewed a sample sequence consisting of three different computer-animated human figures, each performing a distinct action. After a brief delay interval, a test figure appeared and performed an action, and participants indicated,

in separate conditions, whether the test agent, the test action, the test agent or the test action, or the test agent-action combination was present in the sample sequence.

For this first experiment, I minimized the number of cues that differentiated the agents. When beginning the study, participants were told that they would be presented with three identical individuals, differentiated by their clothing color, and that each individual would perform a single action. Other than clothing color, no other cues were given that could be used to differentiate the agents from one another in the sample sequence. To demonstrate that any resulting estimates of storage capacity accurately reflect limitations in VWM with no significant contribution from verbal memory, participants performed a concurrent articulatory suppression task that inhibits the use of verbal coding in memory tasks (Besner, Davies, & Daniels, 1981).

2.1. Method

2.1.1. Participants

Ten male and female participants between the ages of 16 and 30 with normal or corrected-to-normal vision participated to receive credit toward a course requirement or for monetary payment. Informed consent was obtained.

2.1.2. Design

On each trial, observers viewed a sample sequence consisting of three different computer-animated human figures, each performing a distinct action. After a brief delay interval, a test figure appeared and performed an action, and participants indicated whether the relevant test display information was present in the sample sequence, with respect to the following conditions:

- (1) *Agents only*: Participants were told that only the agents could change and to remember only the identities of the agents. On different trials, the test display presented an agent that was not present in the sample sequence; the new agent performed an action that one of the agents from the sample sequence had performed.
- (2) *Actions only*: Participants were told that only the actions could change and to remember only the actions. On different trials, the test display presented an action that was not present in the sample sequence; this action was performed by one of the agents from the sample sequence.
- (3) *Either agent or action*: Participants were told that either the agents or the actions could change and to remember both the identities of the agents and the actions. On 50% of the different trials, the test display presented an agent that was not present in the sample sequence. On the other 50% of the different trials, the test display presented an action that was not present in the sample sequence.
- (4) *Binding agents and actions*: The test display always consisted of an agent and an action that were present in the sample sequence. However, on different trials, the test agent performed an action that had been performed by one of the other agents in the sample sequence. Participants were told to treat such changes as different.

Participants received 50 trials in each condition, preceded by 6 practice trials. The order of conditions was counterbalanced across participants.

2.1.3. 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 sample sequence. The sample sequences consisted of three different sequentially presented human figures each performing a different action. The figures subtended 10.5° (height) \times 4° (width) in the center of a video monitor with a black background. The figures were defined by their clothing color, which was picked at random without replacement from a set of seven highly discriminable colors (red, orange, yellow, green, blue, white, and purple). After appearing, each figure performed an action, which lasted 500 ms and was followed by 500 ms of no motion. Then, the figure was replaced with a new figure (see Fig. 1). The actions were selected at random without replacement from a set of seven highly discriminable actions: forearm curl,

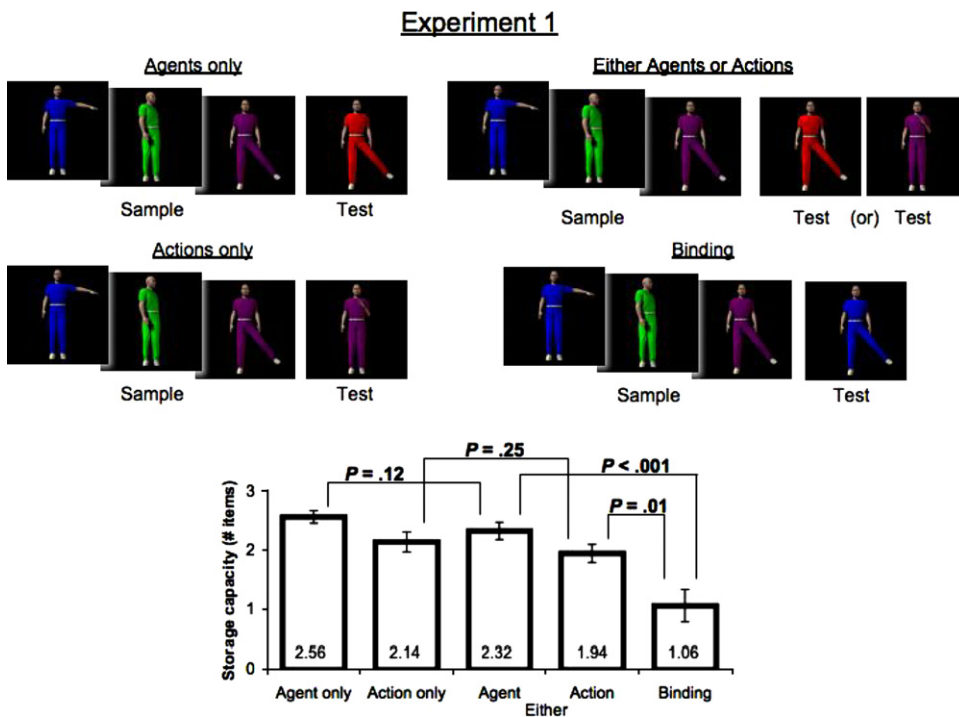


Fig. 1. Example trials of the action sequences and performance for Experiment 1. (Top) Schematic illustration of a trial from each of the four conditions. All examples depict different trials. (Bottom) Performance for each condition, as well as p -values denoting the statistical differences between conditions and trial types as computed through paired-samples t -tests. Error bars denote standard error.

arm raise, head turn, body turn, hand grasp, knee raise, and leg raise. The figures performed the actions on the left side of their bodies.

The sample sequence was followed, after a 500-ms delay interval, by a 1000-ms presentation of the word “test”, followed by the presentation of the test display, which consisted of a single agent performing an action. Participants were required to make a response to the test array, indicating whether the relevant test display information was present in the sample sequence.

2.2. Results and discussion

For the statistical analyses, the data were converted into capacity estimates by using the formula developed by Cowan, 2001 (see also Pashler, 1988). 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 to one of the items on k/n trials. This approach takes into consideration the effects of guessing, by factoring in the false alarm rate ($F = \text{false alarms}/(\text{false alarms} + \text{correct rejections})$) and the observed hit rate ($H = \text{hits}/(\text{hits} + \text{misses})$). The formula is defined as $k = n \times (H - F)$. In the *either* conditions, the average F value of the “same” trials was used for the separate statistical analyses of the agent trials and the action 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, $F(3,27) = 17.42, p < .001$. Post-hoc analyses revealed the pattern of results shown in Fig. 1. Memory capacity in the *actions-only* condition (2.14 actions) was significantly worse than in the *agents-only* condition (2.56 agents). Memory capacity in the *either* condition for the agent (2.32 agents) and action (1.94 actions) information was nearly identical to the memory capacity in the *agents-only* and *actions-only* conditions, respectively. Thus, memory for agents does not interfere with memory for actions, and memory for actions does not interfere with memory for agents. However, memory capacity was significantly lower in the *binding* condition (1.06 integrated representations), when participants needed to remember which agent performed which action, than in the *either* condition, when participants needed to remember the same number of agents and actions but in a non-integrated form, $t(9) = 4.33, p = .002$. Thus, despite the fact that performance in the *either* condition shows that it is possible to retain information about multiple agents and actions simultaneously, these types of information are not maintained together. Rather, they are stored separately in VWM.

Experiment 1 shows that information about agents and actions are stored separately in VWM.¹ These results converge with the findings from Wood (2007) that show that working memory capacity for action and object information are indepen-

¹ A defining characteristic of working memory systems is that there is a severe limit on the amount of information that can be retained at once. Given that in the present study it was possible to retain information about only 2–3 agents and 2–3 actions at one time, performance most likely depended on working memory systems. However, the present results cannot rule out the possibility that long-term memory also played some non-negligible role in this task (see Hollingworth, 2004).

dent, and thus provide further evidence that the visual system consists of separate systems for retaining action information and object/agent information. This raises the question of whether working memory can maintain integrated agent-action representations and, if so, on the basis of what specific visual information. To test this, I conducted three additional experiments that varied the types of visual cues that differentiated the agents from one another.

3. Experiments 2–4

Experiments 2–4 investigated how the visual system binds together agent identity and action information. Three types of cues were inserted into the sample sequences: the agents occupied distinct spatial positions (Experiment 2), a subtle 50-ms gap was inserted between the disappearance of one agent and the appearance of the following agent (Experiment 3), or the agents differed from one another by a variety of physical features (Experiment 4).

3.1. Methods

Ten different participants participated in each experiment. The methods were identical to Experiment 1, except in the following ways:

Experiment 2: The sample agents and the test agent appeared in distinct, non-overlapping locations on the screen. The sample agents appeared on the top (offset 6° from center), left (offset 6° from center), and bottom (offset 6° from center) portions of the screen. The test agent appeared on the right portion of the screen (offset 6° from center) (see Fig. 2).

Experiment 3: A subtle 50-ms gap was inserted between the disappearance of one agent and the appearance of the next agent (see Fig. 2).

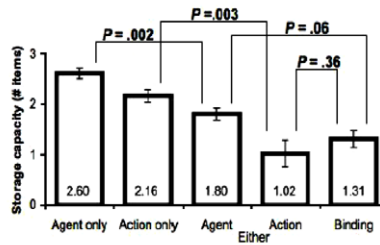
Experiment 4: The agents differed from one another by a variety of physical features (i.e., clothing type, clothing color, gender, age, and facial characteristics). Seven different figures were used (see Fig. 2). Each figure subtended 10.5° (height) × 4° (width) in the center of the screen.

3.2. Results and discussion

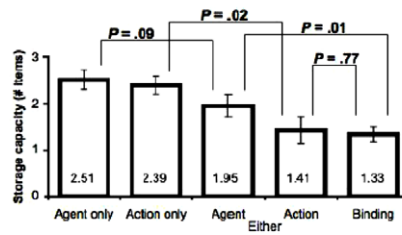
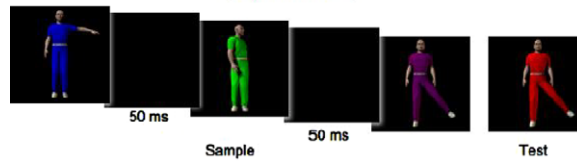
Results were nearly identical for all three experiments. The repeated measures ANOVAs revealed significant main effects of condition for all experiments (Experiment 2: $F(3,27) = 17.87$, $p < .001$; Experiment 3: $F(3,27) = 11.62$, $p < .001$; Experiment 4: $F(3,27) = 15.21$, $p < .001$). Post-hoc analyses (see Fig. 2) provide evidence for two main conclusions, obtained across all three experiments. First, memory

Fig. 2. Example trials of the action sequences and performance for Experiments 2–4. (Top) Schematic illustration of a trial from the *agents-only* condition. All examples depict different trials. (Bottom) Performance for 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.

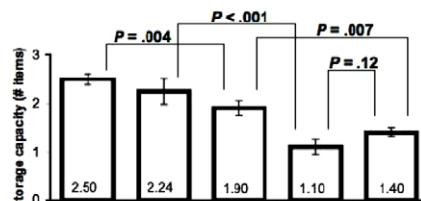
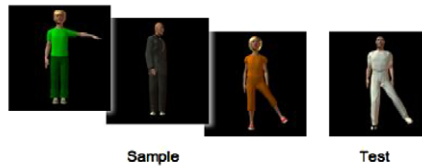
Experiment 2



Experiment 3



Experiment 4



capacity was nearly identical in the *either* and *binding* conditions (Experiment 2: $t(9) = 0.44$, $p = .67$; Experiment 3: $t(9) = 1.63$, $p = .14$; Experiment 4: $t(9) = 0.70$, $p = .50$), with performance in the *binding* condition limited by the type of information that had the lowest memory capacity from the *either* condition (see Fig. 2). This indicates that the visual system used all three types of cues to bind together agent and action information in working memory. Second, memory capacity for both the agent and the action information in the *either* conditions was significantly or nearly significantly lower than memory capacity in the respective *agents-only* and *actions-only* conditions. This indicates that the binding process is resource-demanding – observers can retain less agent and action information when it is bound together.²

For all three experiments, memory capacity in the *agents-only* and *actions-only* conditions was nearly identical to memory capacity for the *agents-only* and *actions-only* conditions from Experiment 1 (all $ps > .25$). This indicates that the impaired memory in Experiments 2–4 for the *either* conditions did not result from increased difficulty in the perceptual analysis of the sequences; rather, it occurred when it was necessary to maintain agent and action information simultaneously. Moreover, participants were able to retain a greater number of bound representations in Experiments 2–4 compared to Experiment 1.

4. Statistical power

Experiment 1 revealed null effects of memory capacity between the *agents-only* and *actions-only* conditions and the respective agent and action trials in the *either* condition. To demonstrate that Experiment 1 had sufficient power to detect these effects of memory capacity, I computed 95% within-subject confidence intervals on the basis of the error terms of the memory capacity factor (Loftus & Masson, 1994). The CI's were $k = \pm .15$ (agent analysis) and $k = \pm .17$ (action analysis), which indicates that Experiment 1 had sufficient power to detect a difference between memory capacity means of .21 or larger ($CI^* \sqrt{2}$) in the agent analysis and .24 or larger in the action analysis. In Experiments 2–4, the average effects of memory capacity between the *agents-only* and *actions-only* conditions and the respective agent and action trials in the *either* conditions were .65 for the agent analyses and 1.09 for the action analyses. Therefore, if the effects of memory capacity observed in Experiments 2–4 had been present in Experiment 1, Experiment 1 had sufficient power to detect them.

² Three additional experiments tested the influence of multiple cues on agent–action binding, by adding the spatial and temporal cues used in Experiments 2 and 3, the temporal and feature cues used in Experiments 3 and 4, and all three cues. Performance in these experiments, when multiple cues were provided, was nearly identical to performance in Experiments 2–4, when only one cue was provided. This indicates that VWM does not benefit from multiple cues when binding together agent and action information.

5. General discussion

The goal of this study was to investigate how the brain maintains visual information about the identities of agents and their actions. Experiment 1 shows that working memory maintains agent and action information in separate memory stores (see also Wood, 2007). Experiments 2–4 show that agent and action information are bound together when the appropriate cues are present in the visual input. However, when agent and action information are bound into integrated agent-action units, VWM can maintain less overall information compared to when this information is not fully bound together, as in Experiment 1.

This pattern of data suggests that the computations supporting the maintenance of information and the computations supporting the binding of information draw upon a common capacity-limited resource, creating a trade-off between the amount of information that can be maintained and whether the information is bound together. How does the visual system allocate resources to these two competing computations? The results suggest that this allocation process is largely determined by the presence of specific cues in the visual input. In Experiment 1, when clothing color differentiated the agents, VWM retained more information in an unbound form, as evidenced by the difference in memory capacity between the *either* and *binding* conditions. Conversely, in Experiments 2–4, when additional cues differentiated the agents, VWM retained less information in a bound form, as evidenced by the nearly identical memory capacities in the *either* and *binding* conditions, both of which were lower than the memory capacities in the *agents-only* and *actions-only* conditions. Note that only the *binding* condition required participants to maintain integrated agent-action representation; in the *either* condition, participants could succeed by remembering the agent and action information separately. Thus, given the trade-off between the amount of information that can be maintained and whether that information is bound together, the best strategy would have been to maintain unbound representations in the *either* condition. Nevertheless, in Experiments 2–4, memory capacity was nearly identical across both the *either* and *binding* conditions. This suggests that the extent to which agent and action information are bound into more integrated units depends largely on the presence of specific cues in the visual input.

What is the relation between the three cues used in Experiments 2–4 with respect to agent-action binding? At the start of Experiment 1, participants were told that they would be presented with three identical individuals, each performing a single action. However, one possibility is that although human adults are free to consider almost anything as an agent given enough time and leisure, the definition of an agent may be well defined at earlier levels of visual analysis. Thus, the color changes that differentiated the agents in Experiment 1 may have been insufficient for the visual system to individuate the agents. Instead, the sequences may have been treated as a single agent performing actions while his clothing changed color, and therefore, fewer resources were allocated to computations for binding agent and action information. The visual system may use additional information, such as the cues presented in Experiments 2–4, as indicators of distinct agents, in which case more

resources are allocated to computations for binding. This proposal is consistent with studies showing that subtle cues change the type of computations performed over a given input. Pena and colleagues, for example, showed that although listeners are able to segment a continuous speech stream on the basis of statistical information, they are unable to extract the structural regularities included in the stream. However, when a 25-ms silent gap was inserted after each word in the speech stream, listeners were able to extract the structural information (Pena, Bonatti, Nespor, & Mehler, 2002). The present method might then be used as a tool to study how the visual system individuates agents from the continuous stream of information it receives through retinal input.

These results have important implications for more general theories of working memory. Currently, it is believed that working memory consists of separate subsystems for retaining verbal information, object information, and location information (Courtney, Petit, Maisog, Ungerleider, & Haxby, 1998; Darling, Della Sala, Logie, & Cantagallo, 2006; Johnson, Hollingworth, & Luck, 2008; Logie, 1995; Luck & Vogel, 1997). The existence of these distinct memory stores has been hypothesized due in large part to the fact that memory capacity is independent for these different types of information; in particular, in studies contrasting visual and verbal stimuli (e.g., Luck & Vogel, 1997; Scarborough, 1972), and in studies contrasting visual and spatial stimuli (e.g., Lee & Chun, 2001). The present study, along with the recent findings obtained by Wood (2007) that show that memory capacity is independent for action and object information and for action and location information, suggests that working memory also consists of an additional subsystem for retaining information about actions (and perhaps other types of events as well). Thus, in order to integrate different types of information that are stored in separate subsystems (e.g., in order to construct an integrated agent-action representation that presumably consists of information from the object subsystem and from the action subsystem), working memory needs to recruit an additional binding process to maintain the links that hold these different types of information together in memory after an event has been observed.

In sum, visual memory consists of parallel memory stores for maintaining information about the identities of agents and their actions. The extent to which agent and action information are bound into more integrated units depends largely on the presence of specific cues in the visual input, which the visual system may use as indicators of distinct agents. This study begins to reveal how the brain organizes the continuous stream of information it receives as input into the agent and action units that serve as the building blocks for many socio-cognitive capacities.

Acknowledgments

I thank Jonathan Beier and Andrew Hollingworth for helpful comments on the manuscript. This work was supported by grants from Harvard University and from a National Institute of Health (NIH) National Research Service Award (NRSA) to

J.N.W. (Grant F31MH075298). Its contents are solely the responsibility of the author and do not necessarily reflect the official views of NIH.

References

- Besner, D., Davies, J., & Daniels, S. (1981). Reading for meaning: The effects of concurrent articulation. *Quarterly Journal of Experimental Psychology*, *33A*, 415–437.
- Courtney, S. M., Petit, L., Maisog, J. M., Ungerleider, L. G., & Haxby, J. V. (1998). An area specialized for spatial working memory in human frontal cortex. *Science*, *279*(5355), 1347–1351.
- Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*, 87–114.
- Darling, S., Della Sala, S., Logie, R. H., & Cantagallo, A. (2006). Neuropsychological evidence for separating components of visuo-spatial working memory. *Journal of Neurology*, *253*, 176–180.
- Downing, P. E., Jiang, Y., Shuman, M., & Kanwisher, N. (2001). A cortical area selective for visual processing of the human body. *Science*, *28*(293), 2470–2473.
- Hollingworth, A. (2004). Constructing visual representations of natural scenes: The roles of short- and long-term visual memory. *Journal of Experimental Psychology: Human Perception & Performance*, *30*(3), 519–537.
- Johnson, J. S., Hollingworth, A., & Luck, S. J. (2008). The role of attention in the maintenance of feature bindings in visual short-term memory. *Journal of Experimental Psychology: Human Perception & Performance*, *34*(1), 41–55.
- Lee, D., & Chun, M. M. (2001). What are the units of visual short-term memory: Objects or spatial locations? *Perception & Psychophysics*, *63*, 253–257.
- Loftus, G. R., & Masson, M. E. J. (1994). Using confidence intervals in within-subjects designs. *Psychonomic Bulletin & Review*, *1*, 476–490.
- Logie, R. H. (1995). *Visuo-spatial working memory*. Hove, NJ: Erlbaum.
- Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*, 279–281.
- O’Craven, K. M., Downing, P. E., & Kanwisher, N. (1999). fMRI evidence for objects as the units of attentional selection. *Nature*, *401*(6753), 584–587.
- Pashler, H. (1988). Familiarity and change detection. *Perception & Psychophysics*, *44*, 369–378.
- Peelen, M., Wiggett, A., & Downing, P. (2006). Patterns of firm activity dissociate overlapping functional brain areas that respond to biological motion. *Neuron*, *49*, 815–822.
- Pena, M., Bonatti, L. L., Nespors, M., & Mehler, J. (2002). Signal-driven computations in speech processing. *Science*, *298*(5593), 604–607.
- Phillips, W. A. (1974). On the distinction between sensory storage and short-term memory. *Perception & Psychophysics*, *16*, 283–290.
- Ruby, P., & Decety, J. (2001). Effect of subjective perspective taking during simulation of action: A PET investigation of agency. *Nature Neuroscience*, *4*, 546–550.
- Scarborough, D. L. (1972). Memory for brief visual displays of symbols. *Cognitive Psychology*, *3*, 408–429.
- Urgesi, C., Candidi, M., Ionta, S., & Aglioti, S. M. (2006). Representations of body identity and body actions in extrastriate body area and ventral premotor cortex. *Nature Neuroscience*, *10*(1), 30–31.
- Wheeler, M. E., & Treisman, 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*(4), 639–652.