Effects of Action Video Game Training on Visual Working Memory

Kara J. Blacker  
Johns Hopkins University

Kim M. Curby  
Macquarie University

Elizabeth Klobusicky and Jason M. Chein  
Temple University

The ability to maintain task-relevant visual information over a brief delay after direct visual input has been removed (i.e., visual working memory) is critical for learning new skills, solving novel tasks, and acquiring new knowledge (e.g., Alloway, Gathercole, & Elliott, 2010; Alloway, Gathercole, Willis, & Adams, 2004; Gathercole & Pickering, 2000; Logie, 2011). Visual working memory (VWM) is the fundamental process that allows us to sustain attended information across saccades and other visual interruptions, to compare objects or scenes based on visual features, and to navigate the visual world. Because VWM is crucial to so many basic actions and processes that guide behavior, research has begun to focus on potential ways to improve this essential system.

The current study focuses on a novel approach to training VWM through the use of action video game play. This approach is seeded in two main areas of literature. First, there is a vast literature regarding individual differences in VWM and how these individual differences are linked to various aspects of visual attention. Second, in the past decade, an emerging body of research has supported the notion that action video game experience enhances a range of perceptual and attentional skills. As will be discussed below, considering these two areas of research together suggests that action video game training may be an exciting and effective new method for training VWM.

One important feature of VWM, especially with regards to training, is its fundamental capacity limitation. Although there is much debate surrounding the nature of the capacity limitation, VWM capacity has commonly been estimated at approximately three or four items (Alvarez & Cavanaugh, 2004; Cowan, 2001; Luck & Vogel, 1997). Despite consistency in the literature about the average capacity of VWM, it is also well documented that VWM capacity varies widely among individuals (Astle & Scerif, 2011; Cowan et al., 2005; Cusack, Lehmann, Veldsman, & Mitchell, 2009; Vogel, McCullough, & Machizawa, 2005). One relevant implication of these known individual differences in VWM is the prospect of using this information to inform the potential remediation of VWM. Considering how crucial VWM is to guiding behavior, investigating how various factors influence individual differences in VWM...
capacity and examining the potential to enhance VWM through training is an important endeavor.

One topic that pervades the literature on individual differences in VWM capacity is the link between VWM and visual attention. Visual attention allows us to select a location or set of locations containing relevant perceptual information within a visual scene. Visual attention and VWM have been characterized as overlapping and interactive processes (e.g., Astle & Searif, 2011; Awh, Vogel, & Oh, 2006; Bundesen, Habelkost, & Kyllingsbaek, 2011; Downing, 2000). Furthermore, there is an extensive body of literature documenting a relationship between individual variability in VWM capacity and visual attention, which we briefly review below. Notably, these established links between attention and VWM may provide a guide for what type of training could most effectively enhance VWM.

Previous research has focused on two particular facets of VWM with regards to attention: the quantity and the quality of information stored in VWM. Interestingly, the literature on the interaction between visual attention and VWM parallels this quantity versus quality distinction. First, there is a strong link between one’s ability to orient attention toward to-be-remembered objects and the quality or precision with which information is encoded and maintained in VWM (e.g., Machizawa & Driver, 2011). Several studies have found that focusing or orienting attention onto a spatial location increases the probability that the information at the oriented location or a spatially nearby location will be encoded into VWM (Jiang, Olson, & Chun, 2000; Linke, Vincente-Grabovetsky, Mitchell, & Cusack, 2011; Schmidt, Vogel, Woodman, & Luck, 2002; Woodman, Vecera, & Luck, 2003). Particularly relevant to training is that high VWM capacity individuals seem to utilize their orienting attention more efficiently than their low capacity counterparts by spatially orienting their attention to groups of objects or locations to be encoded into VWM (Jiang, Chun, & Olson, 2004; Linke et al., 2011). Further, orienting attention to items being maintained in VWM appears to be crucial in protecting and preserving the integrity of information being currently stored in VWM once external input is removed (Griffin & Nobre, 2003; Matsuoka, Luck, & Vecera, 2007; Murray, Nobre, Clark, Cravo, & Stokes, 2013).

Second, selective attention, the ability to attend to task-relevant information while filtering out task-irrelevant information, appears to be very closely related to the quantity of information stored in VWM (e.g., Vogel et al., 2005). Selective attention influences the encoding and maintenance of information in VWM by filtering out task-irrelevant information, thereby reducing the load on an extremely capacity-limited system (Kuo, Stokes, & Nobre, 2012). Specifically, high VWM capacity individuals are more effective at representing only the relevant items in a task than are low capacity individuals, who tend to also encode and maintain information about irrelevant items present in a memory display, thus wasting precious VWM capacity on irrelevant information (Rutman, Clapp, Chadick, & Gazzaley, 2010; Vogel et al., 2005). These two lines of research suggest that training that targets the quality and/or the quantity of information in VWM could yield benefits.

Previous work has shown that category-specific enhancement of VWM is possible with extensive experience. Specifically, Curby and Gauthier (2007) found a VWM advantage for faces compared with other objects. This VWM advantage is attributed to individuals’ expertise with faces. A second study using nonface objects of expertise, cars, as the to-be-remembered items found a similar VWM advantage thereby bolstering the perceptual expertise account of this advantage (Curby, Glazek, & Gauthier, 2009). Considering the inherent visual complexity of faces and cars, perceptual expertise may allow individuals to decrease the information-load of each to-be-remembered item by creating more effective representations, thereby allowing better utilization of the limited VWM system (Curby et al., 2009). A VWM advantage for objects of expertise represents a category-specific benefit, leaving open the question of whether more general improvements to VWM functioning are possible.

Action video game players, who extensively interact with the complex visual environments inherent in action video games, provide an example of extensive experience affording more general visual cognitive enhancements. Studies have shown that prolonged experience playing action video games (e.g., first-person shooter games) enhances the spatial distribution of attention (Green & Bavelier, 2003, 2006a), improves the temporal resolution of attention (Green & Bavelier, 2003), is associated with reduced attentional capture (Chisholm, Hickey, Theeuwes, & Kingstone, 2010; Chisholm & Kingstone, 2012), is associated with improved selective attention (Bavelier, Achtman, Mani, & Focker, 2012; Karle, Watter, & Shedden, 2010), increases visual speed of processing (Dye, Green, & Bavelier, 2009), enhances contrast sensitivity (Li, Polat, Makous, & Bavelier, 2009), improves mental rotation performance (Feng, Spence, & Pratt, 2007), and influences additional facets of visual cognition (for a review see, Achtman, Green, & Bavelier, 2008; also, for a recent meta-analysis see, Powers, Brooks, Aldrich, Palladino, & Alfieri, 2013). Specifically, two studies have previously shown a VWM advantage for action video game players (AVGPs) compared with nonvideo game players (NVGPs) using a change detection task (Blacker & Curby, 2013; Boot, Kramer, Simons, Fabiani, & Gratton, 2008; but also see, Wilms, Petersen, & Vangkilde, 2013). Further evidence suggests that AVGPs use more efficient search strategies in a scene change detection task (Clark, Fleck, & Mitroff, 2011), show superior performance on an enumeration task (Green & Bavelier, 2006b), and represent memory items with greater precision in a delayed localization task (Sungur & Boduroglu, 2012) compared with NVGPs. However, these cross-sectional findings cannot address the question of a causal link between playing action games and enhanced VWM capacity or precision.

Two previous training studies have investigated whether action video game play improves VWM capacity, but the contradictory results do not provide a clear answer (Boot et al., 2008; Oei & Patterson, 2013). Both studies involved having NVGPs train on an action game for approximately 20 hr, but Boot, Kramer, Simons, Fabiani, and Gratton (2008) did not find enhanced VWM capacity after training, whereas Oei and Patterson (2013) did find a VWM improvement in their action group. These contradictory results warrant further investigation into the causal link between action video game play and VWM capacity. Moreover, no study to our knowledge has examined a causal link between action games and enhanced VWM precision/resolution. Sungur and Boduroglu (2012) found evidence for a precision advantage in AVGPs using a cross-sectional design, but a training study is needed to rule out the influence of a selection bias. Therefore, the primary aims of the
current study were to examine the effects of action game training on the quantity and quality of information stored in VWM.

Previous studies examining the effects of action video game play on visual cognition have supported the notion that the enhancements that result from habitual video game play are relatively generalized (i.e., not linked to one specific stimulus category or task). Thus, an interesting question is whether AVGPs’ VWM advantage extends to complex span VWM tasks, which require additional processing and/or manipulation of the to-be-remembered visual information. Here we distinguish between simple span and complex span VWM tasks, where we will use the term “VWM” alone to refer to simple span tasks, such as change detection, also referred to as visual short-term memory (STM) in the literature; and “complex span VWM” to refer to tasks that involve an additional processing component and are thought to reflect higher-order executive skills.

Although these complex span tasks have some overlap with simple span VWM tasks (e.g., change detection), complex span tasks have some unique processes at work as well (Chein, Moore, & Conway, 2011; Unssworth & Engle, 2006, 2007). Complex span tasks differ from traditional simple span or STM tasks in that they require the participant to shift attention away from each successive to-be-remembered stimulus in order to perform some other concurrent task (Chein et al., 2011; Daneman & Carpenter, 1980; Turner & Engle, 1989). Although much research has been dedicated to examining the possibility of improving working memory capacity through training (for a review, see Morrison & Chein, 2011), video game training constitutes a novel method of training compared with traditional working memory training paradigms. Thus, it is unclear whether action video game training will yield any enhancements to complex span VWM.

Recently, concerns have been raised about the utility and validity of video game training studies (Boot, Blakely, & Simons, 2011; Boot & Simons, 2012; but also see Green, Strobach, & Schubert, 2013). Specifically, Boot, Blakely, and Simons (2011) state two important methodological concerns with previous training studies: inadequate baseline for transfer effects and differential placebo effects. Boot et al. (2011) argue that many training studies that show video game training improvements do so in the absence of retest effects for the control group; however, others argue that for many tasks, like those used in the action video game training literature, a single pre-test session is not typically sufficient to drive significant improvements on a single post-test session (Green et al., 2013). However, the idea that a single exposure to a task is insufficient to produce a retest/practice effect does not eliminate Boot et al.’s (2011) second concern that lower expectations for improvement in the control group may suppress any possibility of a retest effect and that positive transfer effects for the action group may simply represent a practice effect for that group. In other words, Boot et al. (2011) suggest that despite the use of a control group, differential placebo effects emerge that may contaminate the training results. Specifically, they argue that a group trained on an action game will expect a different outcome with regards to the assessment tasks compared to those trained on a control game of another genre. This possibility was recently tested by exposing participants to videos of action or nonaction video games and then assessing their expectations regarding how training with these games would be likely to influence performance on a variety of tasks (Boot, Simons, Stothart, & Stutts, 2013). Consistent with the concern raised by Boot et al. (2011), expectations for improvement did indeed differ according to game genre. However, to date, no study has employed a measure of expectations within the context of a training study comparing action and nonaction games. We sought to address these concerns by attaining measures of motivation, engagement, and expectations throughout our training and assessment procedures.

In sum, the current study sought to address the following questions: (a) Is there a causal link between action video games and enhanced VWM capacity? (b) Does action video game play influence the precision with which information is encoded into VWM? (c) Does action video game play enhance performance on a complex span VWM task? and (d) Can motivation, engagement, and/or expectations better account for any training-related findings?

Method

Participants

Thirty-nine male undergraduates with normal or corrected-to-normal vision enrolled in the study. Participants were recruited through an online participant pool and/or through study advertisements and received monetary compensation. All participants were tested for normal color vision prior to enrolling in the study. Five participants dropped out of the study before completion. The remaining 34 participants who completed the study had a mean age of 20.53 years ($SD = 2.57$). Participants were randomly assigned to the action or control group. The 17 action group participants had a mean age of 20.41 years ($SD = 3.04$) and the control group participants had a mean age of 20.65 years ($SD = 2.09$). Participants were unaware of the alternate training condition (i.e., control game participants did not know that other participants were playing an action game and vice versa).

Prior to study enrollment, participants completed a video game experience questionnaire. Participants reported the average number of hours per week that they spent playing various genres of video games in the past year. The genres included: action, fighting, strategy, fantasy, sports, and others. Following previous studies, participants were included in the study if they reported not having played action games in the past year (Green & Bavelier, 2003, 2006a, 2007; Green, Pouget, & Bavelier, 2010; Li et al., 2009). Some participants reported experience with other genres of games in the past year; the action group reported a mean of 1.55 hr/week ($SD = 1.53$) of other genres of games and the control group reported a mean of 1.63 hr/week ($SD = 2.52$).

Due to the documented high correlation between measures of fluid intelligence and VWM capacity (Cowan et al., 2005; Cusack et al., 2009; Fukuda, Vogel, Mayr, & Awh, 2010), participants were tested on the Ravens Progressive Matrices (RPM; Raven, 1990), a standardized nonverbal measure of fluid intelligence, as part of the pre-training assessments. On a split-half version of RPM, participants in the action group scored similarly ($M = 12.59$, $SD = 2.72$) to the control group ($M = 11.88, SD = 3.24$) with no statistical difference between the groups, $t(32) = .69$, $p = .5$. Therefore, the random assignment to the two training groups was.

1 Participants’ video game genre classifications were verified to ensure accurate estimates for each genre of game.
effective, as each group appeared to have a similar distribution of age, nonaction video game experience, and intellectual ability as measured by the RPM.

Video Game Training

The action game training group played Call of Duty®: Modern Warfare® 3 (2011, ActiVision Publishing Inc.) and Call of Duty®: Black Ops® (2010, ActiVision Publishing Inc.). These games were chosen to be similar to those played by our AVGPs in a previous study (Blacker & Curby, 2013) and because of the first-person point of view. Participants played the games in single-player, campaign mode in which the player assumes the role of various characters and attempts to complete different missions. In order to assess game performance and improvement, participants performed two rounds of a “Special Ops” mission, which contains never-ending waves of enemies. This Special Ops mission was performed after 1 hr of training (to serve as a baseline) and after 28 hr of training. These missions provided statistics on shooting accuracy, kills, headshots, amount of time survived, and a composite score, which takes into account all of these variables. Training gains were quantified by examining the composite score improvement from 1 hr to 28 hr of training.

The control game group played The Sims™ 3 (2009, Electronic Arts Inc.). The Sims™ 3 is a simulation-style strategy game, wherein the player takes complete control of the life of a character, which involves everyday activities (eating, bathing, etc.), going to work, managing relationships with other characters, getting married, having and raising children, and eventually growing old and dying. As characters are added to the household, the player takes control of those characters as well. Game performance was measured by tracking: (a) number of characters controlled, (b) number of relationships maintained, and (c) lifetime happiness points, which are acquired throughout the game.

Based on previous training studies, each group trained on their respective game for 30 hours over a duration of 30 days (Green & Bavelier, 2006a, 2007). The groups played their respective games on a 21.5” ViewSonic 1080p HDMI monitor. The action group participants completed the 30 hr of video game training in an average of 27.18 days ($SD = 4.17$) and the control group completed the 30 hr of training in an average of 27.76 days ($SD = 4.40$).

Assessment Battery

All assessments were conducted pre- and posttraining. To avoid order effects for the assessment battery, two different fixed orders were used and each participant was randomly assigned to one of the two orders. The two task orders were evenly distributed between the two groups. All assessment tasks were presented on a 21.5” ViewSonic 1080p HDMI monitor at a viewing distance of 60-cm, with the exception of the color wheel task (details described below).

VWM Capacity Task: Change Detection

Stimuli. Stimuli consisted of colored squares ($1.0° \times 1.0°$), which were displayed on a gray background. The color of each square was chosen randomly without replacement from a set of seven: red, green, yellow, blue, black, white, and purple. Each square was located within an invisible $4 \times 4$ grid subtending a visual angle of $8.5° \times 8.5°$.

Procedure. To measure VWM capacity, a change detection paradigm was used (e.g., Luck & Vogel, 1997). After 500 ms of fixation, each trial consisted of a memory array of four, five, or six colored squares, displayed for 168 ms, followed by a 1,200-ms retention interval and then a 3,000-ms test array (Figure 1a). The memory and test arrays were identical with the exception that on half of the trials, the color of one of the squares was different between the two arrays. During “change” trials, the new color in the test array was selected at random from the other possible colors not shown in the memory array. Participants indicated whether the two arrays were the same or different by a key press. Accuracy was stressed rather than speed and participants received visual feedback after each trial indicating whether their responses were correct or incorrect. Participants performed 32 trials for each set size. The different set sizes were randomly presented in four blocks of 24 trials for a total of 96 trials.

VWM Precision Task: Color Wheel

Stimuli. Stimuli were colored squares ($2° \times 2°$) presented on a gray background. The colors were selected from a master set of 180 evenly distributed and isoluminant hues on a circle in the perceptually homogenous Commission Internationale de l’Eclairage Lab color space (Bays, Catalao, & Husain, 2009; Bays & Husain, 2008; Zhang & Luck, 2008, 2011). Stimuli were presented on a 17” CRT monitor and viewed at a distance of 33 cm.

Procedure. Similar to that in Zhang and Luck (2008, 2011), trials began with a sample array of one, four, or six colored squares displayed for 500 ms. After a 1,200-ms retention interval, participants were shown a test array. The test array contained outlined squares in the same spatial locations as the filled colored squares appeared in the sample array. Surrounding the test array was a color wheel, which contained all 180 possible colors (Figure 1b). Memory array items were always separated by at least two discrete values on the color wheel. One square in the sample array had a thicker outline than the others indicating that it was the probed square. Participants were instructed to use a mouse click to indicate, as precisely as possible, the color of the probed square on the surrounding color wheel. The test array remained on the screen until the participant responded. Participants completed 100 trials for each set size for a total of 300 trials. The different set sizes were randomly presented in three blocks of 100 trials with a 2.5-min break in between blocks.

Data. Recall error was calculated for each trial by measuring the difference, in radians, between the location of the response and the true location of the probed color, where 0 radians would indicate a perfect match to the target color. To avoid an overestimation of error caused by trials where participants randomly guessed (Bays et al., 2009; Sungur & Boduroglu, 2012), any trial where a response was not within the correct half of the circle as the target (i.e., error > $\pi/2$) was excluded.

We chose to use raw error values as our dependent measure of interest here in order to allow for easy comparison to the previous cross-sectional study by Sungur and Boduroglu (2012) demonstrating less error for AVGPs compared with NVGPs on the color wheel task. However, for consistency with the larger VWM pre-
cision literature we have included analyses of precision and mixture modeling results, as detailed in Bays et al. (2009), in the supplementary online material.

Complex Span VWM Task: Symmetry Span

The symmetry span task is a complex span task that has been previously shown to be sensitive to individual differences in VWM capacity estimates (Kane et al., 2004) and is distinct from simple span measures like change detection described above (Unsworth & Engle, 2007).

**Procedure.** Participants recalled sequences of red-square locations within a matrix against a background symmetry judgment task (Figure 1c). In the symmetry judgment task, an 8 × 8 matrix (7° × 7°) was presented, with some squares filled in black, and participants had to decide whether the black-square design was symmetrical along its vertical axis; it was symmetrical approximately half the time. Participants responded via a button press. Following the participant’s response to the symmetry display, a 500-ms delay was followed by a 4 × 4 matrix with a filled red square presented for 650 ms. Immediately following the to-be remembered matrix, either another symmetry display or the recall cue was presented. When presented with the recall cue, participants were asked to click, using a mouse on a blank 4 × 4 grid, the sequence of red-square locations in the order in which they appeared in the preceding display. Sets of two, three, four, or five symmetry–memory matrices were presented per trial. Participants completed three trials per set size for a total of 12 trials.

**Data.** Two scores were calculated separately: an “absolute score” and a “partial score.” The absolute score is the sum of all trials in which all red squares were recalled in the correct location and serial order. The partial score is the sum of red squares recalled in the correct location and serial order, regardless of whether the entire trial was recalled correctly. For example, in a Set Size 3 trial, if two of the three squares were correctly recalled, then two is added to the partial score, but nothing is added to the absolute score because all three items were not correctly recalled. Therefore, the absolute score must always be less than or equal to the partial score. Accuracy was the dependent measure for the symmetry judgment task. Participants with an accuracy level below 80% were excluded because the integrity of the central memory task is dependent on the participant engaging adequately with the symmetry task.

Additional Assessment Tasks

Three additional assessment measures were used and the results are reported in the supplementary online material. These tasks were not central to our key hypotheses of whether action video game training improves VWM and were used as “spacer” tasks to ensure participants did not fatigue on measures of VWM and to allow comparison to a separate cohort of individuals participating
in a working memory training study. The tasks included: (a) an immediate free-recall task used to measure verbal working memory (Craik, 1970); (b) the traditional Stroop task used to measure cognitive control (Stroop, 1935); and (c) a behavioral version of the antisaccade task used as a measure of proactive cognitive control (Guittion, Buchtel, & Douglas, 1985). In an effort to adhere to the methodological suggestions of Boot et al. (2011), we report all assessment tasks, but again these measures were not of central importance to our questions of interest and were used as filler tasks, which is why the results are presented only in the supplementary material.

**Motivation and Engagement Measures**

Prior to each of the assessment tasks, participants were asked to complete a self-report measure of how motivated they were to perform each task using a 0 to 9 scale (0 = not motivated at all, 9 = extremely motivated). After each assessment task, participants were then asked to complete a self-report question about how engaged they had been in the task using a 0 to 9 scale (0 = not engaged at all, 9 = extremely engaged). The same motivation and engagement scales were completed before and after each training session, respectively. These scales were added to assess whether there were any differences in motivation/engagement levels between the two groups during training or the assessment tasks.

**Exit Survey**

Participants completed an exit survey upon completion of the study, which was aimed to evaluate their experience in the study and to determine what their expectations of the purpose and outcome of the study were. The exit survey was administered to participants immediately after they completed the final assessment task. Participants were asked about their previous knowledge of research regarding video game playing, both costs and benefits (i.e., “Before enrolling in or during your enrollment in this study, were you familiar with any research about any potential positive or negative effects of video game playing?”). If participants responded “yes,” they were asked to describe the research they were familiar with in sentence form. Participants were also asked how much they enjoyed the game that they trained on using a Likert-type scale ranging from 1 (I disliked the game very much) to 5 (I enjoyed the game very much).

Finally, for each assessment task participants were asked to rate how they thought their 30 hr of video game training influenced their performance from pre- to posttraining. More specifically, they were presented with the following prompt: “For each of the following assessment tasks that you completed, what effect do you think your video game training had on your performance.”2 They again used a Likert-type scale: 1 (major decrease), 2 (minor decrease), 3 (no change), 4 (minor increase), 5 (major increase). Each of the assessment tasks were then presented using the following descriptions: change detection, “Colored Squares (did the array of colored squares change);” color wheel, “Color Wheel Memory;” and symmetry span, “Red square memory and symmetry judgments.” To alleviate reliance on memory for the tasks, participants were instructed to ask the experimenter, who was present during the exit survey, for further clarification if they were unsure, of which task the description was probing. No participants requested clarification. These data were collected in an effort to address an ongoing concern in the literature that participants in training studies may have different expectations about the effects of playing action versus nonaction video games that could impact their posttraining performance (Boot et al., 2011).

**Results**

**Motivation and Engagement**

To examine whether motivation and engagement levels during training differed between the two groups, independent samples $t$ tests were used to compare average motivation and engagement scores across training sessions. No significant differences in motivation, $t(32) = .73, p = .47$, or engagement, $t(32) = .49, p = .63$, emerged between the action and control groups during training.

To further assess any potential differences in motivation or engagement between groups, 2 (Time: pre-, posttraining) $\times$ 2 (Group: action, control) repeated-measures ANOVAs were performed on motivation and engagement scores for each assessment task separately. For all assessments, no significant main effects or interactions emerged, change detection task: all $ps \geq .28$, color wheel task: all $ps \geq .16$, symmetry span task: all $ps \geq .16$. These results suggest that individuals were not more or less motivated or engaged in the assessment tasks before or after training (i.e., no main effects of time), neither group was more or less motivated or engaged with regard to the assessment tasks (no main effects of group), nor were there any interactions between time and group. Thus, any performance enhancements after training cannot be readily explained by differential task motivation or task engagement.

**Training Performance**

The composite score for the Special Ops mission was used as a measure of training improvement and was shown to be reliable across all attempts (see Table 1). Fifteen of the 17 action participants showed improvement from Hour 1 of training to Hour 28 of training on this Special Ops training assessment. However, two participants did not improve and showed below average advancement through the game on all measures, including large negative composite scores from Hours 1 to 28 of training, and were therefore not included in any further analyses. These two participants were also in the bottom quartile of the group for self-reported motivation scores during both training and assessments.

For The Sims game, the primary measure of performance was the “Lifetime Happiness” points that are acquired throughout the game. Participants accumulated an average of 293,045 ($SD = 168,427$) lifetime happiness points. Participants controlled an average of 3 ($SD = 2$) characters and maintained an average of 61 ($SD = 46$) relationships throughout the 30 hr of training. Based on these training measures, there were no statistical outliers (i.e., performance $> 2 SD$ above or below the group mean) and therefore all control participants were included in the following analyses, unless otherwise specified below.

---

2 Participants answered the same expectation question for each of the additional assessment tasks based on similarly descriptive probes for each.
Table 1
Test–Retest Reliability for Each Measure of Interest

<table>
<thead>
<tr>
<th>Measure</th>
<th>Cronbach’s alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td>Special Ops training performance</td>
<td>0.72</td>
</tr>
<tr>
<td>Change detection</td>
<td>0.60</td>
</tr>
<tr>
<td>Color wheel</td>
<td>0.73</td>
</tr>
<tr>
<td>Symmetry span</td>
<td>0.78</td>
</tr>
</tbody>
</table>

Note. For change detection and color wheel, the Cronbach’s alpha was calculated for each set size and the average across set sizes is presented here. Similarly, for symmetry span, the average of the partial and absolute scores is presented.

Change Detection

Participants were excluded from the analysis if they demonstrated chance-level performance in either the pre- or posttraining assessment task, which resulted in the exclusion of two control group participants. A response time (RT) filter was not used because participants were given a maximum of 2,500 ms to respond and extremely fast RTs (<200 ms) were rare (0.005% of total trials). Analysis of the change detection task targeted differences in training gains for the action and control groups. A 3 (Set Size: 4, 5, 6) × 2 (Group: action, control) × 2 (Time: pre-, posttraining) repeated-measures ANOVA was conducted on accuracy values. Significant main effects of set size, $F(2, 27) = 28.48, p < .001$, $\eta^2_p = .50$, and time, $F(1, 28) = 6.31, p < .05$, $\eta^2_p = .18$, emerged, with greater accuracy at smaller set sizes and at the posttraining assessment time point, respectively. Importantly, there was also a significant Time × Group interaction, $F(1, 28) = 5.17, p < .05$, $\eta^2_p = .16$. To examine this interaction further, paired-samples $t$ tests were tested on accuracy values for each group separately. The action group demonstrated a significant gain in performance from pre- to posttraining, $t(14) = 3.42, p < .01$, whereas the control group did not, $t(14) = .17, p = .87$ (see Figure 2). Finally, the main effect of group and the remaining interactions did not approach significance, $F$s ≤ 1.1, $p$s ≥ .33.

To compare the size of these training effects with reference to capacity estimates, Pashler’s $K$ (set size−[(hit rate − false alarm rate)/(1 − false alarm rate)]) was also calculated (Pashler, 1988; Rouder, Morey, Morey, & Cowan, 2011). Because all presented set sizes were at or above normal capacity limits, maximum $K$-value ($K_{max}$) across all set sizes is considered here. Prior to training, the action group stored an average of 4.0 items ($SD = .93$), whereas after training they stored an average of 4.6 items ($SD = .75$), which represents a significant training gain, $t(14) = 3.44, p < .01$. In comparison, the control group stored an average of 3.98 items ($SD = .71$) before training and an average of 4.14 items ($SD = .95$) after training, which was not a significant gain, $t(14) = .73, p = .48$. In sum, the action group showed significantly more posttraining improvement on the change detection task compared with the control group, which suggests that experience playing action video games enhanced VWM capacity.

Color Wheel

Analysis for the color wheel task targeted group differences in precision, as measured by the degree of response error, before and after training. Trials with extremely fast (<200 ms) or extremely slow (>4,000 ms) RTs were excluded (3.4%). Participants who lost more than 25% of trials due to this RT filter were excluded from additional analyses ($n = 1$). A 3 (Set Size: 1, 4, 6) × 2 (Time: pre-, posttraining) × 2 (Group: action, control) repeated-measures ANOVA was conducted on error values. A main effect of set size emerged, $F(2, 28) = 378.49, p < .001$, $\eta^2_p = .93$, with error increasing as set size increased. No other main effect approached significance, $p$s ≥ .81. The Time × Group interaction did not reach significance, $F(1, 29) = 1.28, p = .27$, $\eta^2_p = .04$, nor did any other two-way interactions, $p$s ≥ .34. However, the three-way, Group × Time × Set Size, interaction approached significance, $F(2, 28) = 3.02, p = .056$, $\eta^2_p = .09$ (see Figure 3).

To explore this marginally significant three-way interaction further, separate 2 (Time) × 2 (Group) repeated-measures ANOVAs were
conducted on each set size separately. For Set Size 1, neither main effect of group or time, nor their interaction approached significance, $p_s \geq .47$. For Set Size 4, neither main effect of group or time was significant, $p_s \geq .73$; but importantly, the Time $\times$ Group interaction did reach significance, $F(1, 29) = 5.49, p < .05$, $\eta^2 = .16$, with the action group showing a larger reduction in error after training, relative to the control group. Finally for Set Size 6, neither main effect of group or time, nor the interaction reached significance, $p_s \geq .42$. These results suggest that the action group showed significant decreases in response error, and therefore greater precision, relative to the control group, but only at Set Size 4.

These results tentatively suggest that there was a significant training effect for the action group on the color wheel task; tentative because the interaction involved numerically decreased performance (more response error) in the control group posttraining, and posttraining performance in the action group that was comparable to that exhibited by the pre-training control group (see Figure 3). Further implications of these results will be elaborated on in the Discussion section.

**Symmetry Span**

Analysis for the symmetry span task targeted differences in training gains in complex span VWM for the action and control groups. The requirement that participants perform above 80% accuracy on the symmetry judgment resulted in the exclusion of data from four participants. An RT filter was not used because this task requires multiple responses per trial. To examine potential training differences between the two groups, 2 (Group: action, control) $\times$ 2 (Time: pre-, posttraining) repeated-measures ANOVAs were conducted on absolute and partial scores separately. For absolute score, a significant main effect of time emerged, $F(1, 26) = 4.61, p < .05$, $\eta^2 = .15$, with performance being greater posttraining. Neither the main effect of group, $F(1, 26) = .04, p = .85$, $\eta^2 = .001$, nor the Time $\times$ Group interaction, $F(1, 26) = .29, p = .59$, $\eta^2 = .01$, approached significance for partial score. For absolute score, the main effects of time, $F(1, 26) = 2.32, p = .14$, $\eta^2 = .08$, and group, $F(1, 26) = .01, p = .90$, $\eta^2 = .001$, did not reach significance, nor did the Time $\times$ Group interaction, $F(1, 26) = .17, p = .69$, $\eta^2 = .01$. As can be seen in Figure 4, action versus nonaction video game training does not appear to have differentially improved complex span VWM performance.

**Exit Survey**

To assess previous knowledge of video game research and expectations of performance changes, the exit survey data was assessed for each training group. Of the 34 participants who completed the study, seven (two action, five control) reported being aware of research about the negative effects of video games and nine participants (five action, four control) reported being aware of research about the positive effects of video games prior to enrolling in the study.

In line with the motivation and engagement results, there was no significant group difference in the degree to which participants enjoyed playing their respective game. The action group reported a mean enjoyment of 4.29 ($SD = .99$), which was similar to the control group’s mean enjoyment score of 3.82 ($SD = 1.42$), $t(32) = 1.12, p = .27$.

Most importantly, after training participants reported whether they expected an increase or decrease in their task performance for each assessment task using a 1–5 Likert-type scale. We tested a 3 (Task) $\times$ 2 (Group) repeated-measures ANOVA on expectation scores. A main effect of task emerged, $F(2, 28) = 11.71, p < .001$, $\eta^2 = .29$, with participants expecting the most improvement on the symmetry span task and the least improvement on the color wheel task. The main effect of group did not reach significance, $F(1, 29) = 3.67, p = .065$, $\eta^2 = .11$. However, it is worth noting that although this main effect of group did trend in favor of higher expectations for the action group, the biggest numerical group difference in expectations was on the symmetry span task, which showed the least evidence for a training-specific effect (see Figure 5). The Task $\times$ Group interaction was also not significant, $p = .85$. As can be seen in Figure 5, for the change detection, $t(32) = .83, p = .41$, color wheel, $t(32) = .82, p = .42$, and symmetry span, $t(31) = 1.68, p = .1$, tasks there was no significant difference between performance expectations between the two groups. These results imply that differential expectations between the two training groups cannot account for the training improvements in VWM capacity or precision described above.

Furthermore, we investigated the sensitivity of our assessment measures to expectation effects by examining potential correlations between expectation ratings and task performance gains. There were no significant correlations between expectations on the color wheel task and performance gains for any of the set sizes, all $p_s \geq .5$. There were also no significant correlations between expectations and performance gains for the absolute or partial score of the symmetry span task, $p_s \geq .5$. These results suggest that both the color wheel and symmetry span tasks were insensitive to expectations. Interestingly, there was a significant correlation between expectations and performance (e.g., average accuracy across set sizes) for the change detection task, $r(30) = .50, p < .01$. However, when the relationship between expectations and performance for the change detection task was examined by training group, it appears that the correlation was significant for the control group alone, $r(15) = .69, p < .01$, but not for the action group alone, $r(15) = .40, p = .14$. However, there is not a significant difference between these two correlations, $p = .28$. 

![Figure 4](image-url)
which may suggest that the change detection task was sensitive to expectations. Despite this possibility, given that there were no overall differences in expectations on the change detection task between the action and control group, differential expectations cannot explain the action groups’ training-related improvement to VWM.

Discussion

VWM is a ubiquitous process that allows us to maintain a stable representation of the visual world in the service of many visual tasks. Despite the constant utility of VWM in our everyday lives, there is a severe capacity limitation to this system. Excitingly, the results presented here suggest the potential malleability of VWM through training. Building on the existing literature on action video games enhancing various aspects of visual attention and performance, the current study examined whether action video game training could enhance VWM capacity, VWM precision, and/or complex span VWM. Individuals who trained on an action game demonstrated significant improvement on a measure of VWM capacity compared with those who trained on a control game. In addition, the action group showed some improvement on a measure of VWM capacity compared with those who trained on a control game. In addition, the action group showed some improvement on a measure of VWM capacity compared with those who trained on a control game. In addition, the action group showed some improvement on a measure of VWM capacity compared with those who trained on a control game. In addition, the action group showed some improvement on a measure of VWM capacity compared with those who trained on a control game. In addition, the action group showed some improvement on a measure of VWM capacity compared with those who trained on a control game. In addition, the action group showed some improvement on a measure of VWM capacity compared with those who trained on a control game. In addition, the action group showed some improvement on a measure of VWM capacity compared with those who trained on a control game. In addition, the action group showed some improvement on a measure of VWM capacity compared with those who trained on a control game. In addition, the action group showed some improvement on a measure of VWM capacity compared with those who trained on a control game. In addition, the action group showed some improvement on a measure of VWM capacity compared with those who trained on a control game. In addition, the action group showed some improvement on a measure of VWM capacity compared with those who trained on a control game. In addition, the action group showed some improvement on a measure of VWM capacity compared with those who trained on a control game. In addition, the action group showed some improvement on a measure of VWM capacity compared with those who trained on a control game. In addition, the action group showed some improvement on a measure of VWM capacity compared with those who trained on a control game. In addition, the action group showed some improvement on a measure of VWM capacity compared with those who trained on a control game. In addition, the action group showed some improvement on a measure of VWM capacity compared with those who trained on a control game. In addition, the action group showed some improvement on a measure of VWM capacity compared with those who trained on a control game. In addition, the action group showed some improvement on a measure of VWM capacity compared with those who trained on a control game. In addition, the action group showed some improvement on a measure of VWM capacity compared with those who trained on a control game. In addition, the action group showed some improvement on a measure of VWM capacity compared with those who trained on a control game. In addition, the action group showed some improvement on a measure of VWM capacity compared with those who trained on a control game. In addition, the action group showed some improvement on a measure of VWM capacity compared with those who trained on a control game. In addition, the action group showed some improvement on a measure of VWM capacity compared with those who trained on a control game. In addition, the action group showed some improvement on a measure of VWM capacity compared with those who trained on a control group. The results from the color wheel task also provide tentative support for enhanced VWM among the action group as a consequence of the training with action video games.

Using a change detection paradigm, the action group demonstrated a significant training-related gain in VWM capacity, whereas the control group did not. This finding is consistent with previous cross-sectional studies, demonstrating a VWM advantage among avid AVGPs (Blacker & Curby, 2013; Boot et al., 2008). Although Boot et al. (2008) did not find a training effect for VWM capacity, the current study used 10 more hr of training (i.e., a 50% increase), which may have been a critical factor in the significant training-related improvement reported here. The current results are consistent with those recently reported by Oei and Patterson (2013). Although Oei and Patterson (2013) did not report a direct comparison of their action groups’ training gains to the other group(s) in their study, here it is demonstrated that action video game training improved change detection accuracy in our action group, but not in our control group. The current results also extend the findings of Oei and Patterson’s (2013) by using a more conventional training display monitor (i.e., a 21.5” computer monitor here compared with the 3.5” inch screen used by Oei & Patterson, 2013). A final methodological difference between the current study and that of Boot et al. (2008) and Oei and Patterson (2013) is the action game of choice utilized. All three studies used fast-paced, first-person shooter action games, which according to Cohen, Green, and Bavelier (2007) are the pertinent factors. We chose the Call of Duty series for the current study because it was the most frequently cited action game among our action gamer group in our previous cross-sectional study (Blacker & Curby, 2013); however, an important aspect of future work will be to better understand the crucial mechanistic features of these action games to provide insight into which games work best for training and why.

In the context of research linking individual differences in VWM capacity with the effectiveness of selective attention (Kuo et al., 2012; Rutman et al., 2010; Vogel et al., 2005), the enhanced VWM performance by our action group is consistent with previous findings of enhanced selective attention after extensive action video game play (Bavelier et al., 2012; Chisholm et al., 2010; Chisholm & Kingstone, 2012). Thus, given that (a) action video game training enhances both VWM and selective attention (Bavelier et al., 2012), and (b) individual differences in VWM capacity can be predicted, in part, by an observer’s ability to selectively attend to task-relevant information, enhancements to selective attention may underlie the VWM advantage observed after action game training. In other words, it may be that action video games directly enhance the VWM system itself or that the improvement occurs indirectly via enhancement of selective attention, which in turn bolsters VWM performance. However, neither the current study nor previous studies have explicitly tested this possibility making it an important endeavor for future research.

The results reported here provide strong support for the responsiveness of VWM to training. More specifically, the change detection results suggest that VWM performance is malleable to some degree, as evidenced by the action group’s enhanced capacity estimates after training compared with those for the control group. The results from the color wheel task also provide tentative support for enhanced VWM among the action group as a consequence of the training with action video games.

Before considering the color wheel results in detail, there is a potential limitation to these findings that warrants discussion. The positive training effects seen for the action group at Set Size 4 in the color wheel task were captured in the presence of negative effects for the control group (i.e., the control group showed more response error after training). We had no a priori expectation that playing the control game for 30 hr would decrease performance on this task. However, one could speculate that the drastically different color palettes that dominated the action and control games may have played a role. Specifically, although the former used subtle color variations (with camouflage), the latter included exaggerated color variations, which could have potentially influenced color judgments on the color wheel task. This account is obviously very
speculative. Nonetheless, previous demonstrations of the plasticity of the neural mechanism underlying color perception even in adulthood, specifically in terms of the impact of chromatic experience suggests that this is an intriguing possibility (e.g., Kwon et al., 2011; Neitz, Carroll, Yamauchi, Neitz, & Williams, 2002). Alternatively, this finding could reflect a Type I error (the 3-way interaction present in the color wheel results was only marginally significant), which limits the conclusions that can be made on the basis of these data. This potential limitation warrants the need for a replication to further test the effects of video game training on VWM precision estimates.

Despite their tentative nature, findings from the color wheel task warrant consideration in the context of the existing literature on action video game benefits. The finding of improved precision after action game training is consistent with previous cross-sectional results revealing that AVGPs store more precise VWM representations (Sungur & Boduroglu, 2012). However, the set size specificity of this finding, with it only occurring when the set size was four items, is open to a number of interpretations, only one of which will be discussed in detail here. One possibility is that observing an improvement in VWM precision after action video game training may depend on the VWM system being optimally driven, and thus it may not appear when the system is not challenged sufficiently (Set Size 1) or is overwhelmed (Set Size 6). Note, the Set Size 4 condition most closely to the functional capacity of the group as a whole, which may support this explanation. The somewhat tentative nature of our findings regarding the impact of video game training on the precision of VWM representations prohibit us from making any strong claims. However, our finding and that of a previous cross-sectional study demonstrating a similar effect (Sungur & Boduroglu, 2012) highlight the need for further work exploring a potential causal link between action video game experience and VWM precision.

The complex span task results reported here argue against the possibility that the VWM advantage observed after action video game training extends to complex span VWM. Although participants appeared to improve on the complex span task from pre- to posttraining, there was no evidence of a training-specific effect. Both groups improved to about the same degree, which suggests this effect may have been due to practice effects with the task itself. Although simple span tasks, like change detection, and complex span tasks, like symmetry span tap into related processes, they involve at least one important distinction: The symmetry span task required the individual to not only store memory items and later recall them, but also to direct attention away from those memory items and perform a secondary processing task during the maintenance period. Although action video game experience appears to enhance the capacity of VWM as measured by change detection, it does not seem to extend to complex span measures. There may be several reasons why this is the case, but two are specifically discussed here. First, it may simply be the case that 30 hr of training is not sufficient to bring about significant changes in complex span task ability. Working memory, as measured by complex span tasks, is often conceptualized as a domain-general process that underlies many other cognitive abilities. Thus, enhancing higher-order executive skills, like those utilized in complex span VWM tasks, may require more extensive training and/or more targeted training (for a review, see Morrison & Chein, 2011). This possibility is consistent with previous studies showing no difference between AVGPs and NVGPs on tasks that tap into executive skills, such as dual-tasks paradigms (e.g., Donohue, James, Eslick, & Mitroff, 2012). Second, action video games require constant encoding and subsequent use of visual information, but it is less clear to what extent complex span VWM abilities, like those tested in the symmetry span task, are utilized in action games. For example, action games are so fast-paced that there is rarely a need or an opportunity to store a piece of visual information and then manipulate it or direct attention away before utilizing the stored information. More commonly, visual information is stored and quickly needed again before many other events or stimuli intervene. Therefore, it may be that action games utilize quick encoding and brief storage of visual information, which is why the action group improved on change detection, but not on the symmetry span task. Perhaps other genres of video games or other types of training, which place higher demands on storing and subsequently manipulating visual information, would yield more productive training effects for complex span VWM.

Of note are potential concerns that have been raised about the utility and validity of video game training studies more generally (Boot et al., 2011). Specifically, Boot et al. (2011) state two important methodological concerns with previous training studies: inadequate baseline for transfer effects and differential placebo effects. The results here for all assessment tasks demonstrated some degree of retest effects for both groups (with the exception of the color wheel task, discussed above), as evidenced by significant main effects of time for all assessment measures and all measures of interest showed acceptable test–retest reliability (see Table 1). In addition, upon the suggestion of Boot et al. (2011), an exit survey was used in which participants rated their expectations for improvement on each assessment task based on their training experience. These data revealed no differences in expectations between the two groups on any of the tasks. This pattern of results suggests that the participants were unable to predict the outcome of the study in any meaningful way. One potential caveat to our expectations assessment is the use of a 5-point Likert-type scale, which may not have had sufficient sensitivity to reveal a possible effect of expectations. To our knowledge, this is the first video game training study to implement a measure of expectations. Thus, there is no standard practice by which to follow and therefore future studies should also investigate other methods of assessing training expectations (Boot et al., 2013). Moreover, there were no differences between the two groups either on their self-reported enjoyment of the game they trained on, or on their self-reported motivation and engagement levels during the training or assessments. Thus, the training-related improvement in the action group seen for VWM capacity cannot be attributed to inadequate baseline measures or differences in motivation, engagement, or performance expectations.

The enhancement of VWM performance after action video game training is relevant to the active debate in the literature regarding what constrains VWM capacity (for reviews see Brady, Konkle, & Alvarez, 2011; Luck & Vogel, 2013). The current study examines the potential dissociation between the quantity of information encoded and maintained in VWM versus the quality of that information. We suggest that action video game training enhances the quantity, but the improvement in quality is less certain given the tentative nature of the color wheel results here. One interesting question for future work is to explore whether improvement in
quantity can also increase quality and vice versa. A training-based approach to this question of quality versus quantity in VWM may aid in the current debate about whether these aspects of VWM are dissociable. Although the data here do not necessarily differentiate between a slots-based model (Luck & Vogel, 1997; Zhang & Luck, 2008, 2011) or a resource-based model (Bays et al., 2009; Gorgoraptis, Catalao, Bays, & Husain, 2011), an important future direction for these competing models will be not only to account for the limiting factors impinging on VWM capacity, but to account for potential increases in the performance of this system after training.

In addition to their potential theoretical implications, the findings reported here also have implications for the remediation of VWM deficits and cognitive training more broadly. Based on these findings, action video game training may represent a unique and engaging platform in which to train one’s capacity-limited VWM system. One practical caveat to the use of action video games as a training method is the typically violent nature of action games (e.g., for meta-analytic reviews on the negative effects of violent video games, see Anderson & Bushman, 2001; Greitemeyer & Mugge, 2014). First-person shooter, action games have been specifically shown to produce these visual cognitive enhancements, above and beyond that of other genres of games (Cohen, Green, & Bavelier, 2007; also see Oei & Patterson, 2013). Therefore, an important next step in this research will be to better understand the mechanisms driving these training-related improvements to one’s visual skills. By parsing out what drives these enhancements, training could become tailored to specific visual skills (e.g., VWM, contrast sensitivity, temporal resolution of attention) or to specific populations (e.g., nonviolent games for children with visual cognitive deficits).

Action video games provide the player with a complex and constantly changing visual environment in which accurate visual memory and acute attentional skills often determine the player’s success or failure in the game. Exposure to these visual environments over an extensive period of time appears to enhance VWM performance. Therefore, action video game training represents a distinctive form of training that may allow individuals to exploit the malleability of this inherently capacity-limited VWM system.

References


Received December 12, 2013
Revision received June 25, 2014
Accepted June 26, 2014