Psychological Research
DOI 10.1007/s00426-015-0648-y

ORIGINAL ARTICLE

Working memory training improves visual short-term memory capacity

Hillary Schwarb · Jayde Nail · Eric H. Schumacher

Received: 17 March 2014 / Accepted: 22 January 2015

Abstract Since antiquity, philosophers, theologians, and scientists have been interested in human memory. However, researchers today are still working to understand the capabilities, boundaries, and architecture. While the storage capabilities of long-term memory are seemingly unlimited (Bahrick, J Exp Psychol 113:1–2, 1984), working memory, or the ability to maintain and manipulate information held in memory, seems to have stringent capacity limits (e.g., Cowan, Behav Brain Sci 24:87–185, 2001). Individual differences, however, do exist and these differences can often predict performance on a wide variety of tasks (cf. Engle What is working-memory capacity? 297–314, 2001). Recently, researchers have promoted the enticing possibility that simple behavioral training can expand the limits of working memory which indeed may also lead to improvements on other cognitive processes as well (cf. Morrison and Chein, Psychol Bull Rev 18:46–60 2011). However, initial investigations across a wide variety of cognitive functions have produced mixed results regarding the transferability of training-related improvements. Across two experiments, the present research focuses on the benefit of working memory training on visual short-term memory capacity—a cognitive process that has received little attention in the training literature. Data reveal training-related improvement of global measures of visual short-term memory as well as of measures of the independent sub-processes that contribute to capacity (Awh et al., Psychol Sci 18(7):622–628, 2007). These results suggest that the ability to inhibit irrelevant information within and between trials is enhanced via *n*-back training allowing for selective improvement on untrained tasks. Additionally, we highlight a potential limitation of the standard adaptive training procedure and propose a modified design to ensure variability in the training environment.

Introduction

The idea that minimal practice with simple cognitive tasks can lead to performance improvements on a wide variety of untrained skills is enticing. In fact, popular culture has embraced this possibility and there are now several commercially available "brain training" programs promising to improve cognition (viz. Owen, Hampshire, Grahn, Stenton, Dajani & Burns, 2010; Shipstead, Redick & Engle, 2010). However, despite the enthusiasm surrounding cognitive training, evidence for the efficacy of such programs is inconsistent (Owen et al., 2010; Melby-Lervåg & Hulme, 2013; Morrison & Chein, 2012). While several studies have provided support for training-related improvements on a variety of untrained tasks (e.g., Chein & Morrison, 2010; Green & Bavelier, 2003, 2006; Jaeggi, Buschkuehl, Jonides & Perrig, 2008; Jaeggi, Studer-Luethi, Buschkuehl, Su, Jonides & Perrig, 2010b), other studies have reported no such improvements on similar and even identical untrained tasks (e.g., Boot, Kramer, Simons, Fabiani & Gratton, 2008; Owen et al., 2010; Redick, Shipstead, Harrison, Hicks, Fried & Hambrick, 2012; Melby-Lervåg & Hulme, 2013).

Popular culture and the scientific community alike have taken a particular interest in working memory (WM)

H. Schwarb (✉)
Beckman Institute, University of Illinois, 405 North Mathews Avenue, Urbana, IL 61801, USA
e-mail: schwarb2@illinois.edu

J. Nail · E. H. Schumacher
School of Psychology, Georgia Institute of Technology, 654 Cherry Street, Atlanta, GA 30332, USA

training because WM appears to be a central component to critical real-world abilities such as fluid intelligence (Gf; e.g., Engle, Kane & Tuholski, 1999; Kane, Hambrick, Tuholski, Wilhelm, Payne & Engle, 2004), mind wandering (Kane, Brown, McVay, Silvia, Myin-Germeys & Kwapil, 2007a), and controlled attention (e.g., Conway, Cowan & Bunting, 2001; Kane & Engle, 2003; Unsworth, Schrock & Engle, 2004) to name a few. Thus, the potential benefit of enhancing WM capacity has seemingly limitless real-world applications.

Although most studies estimate that mean WM capacity is approximately four items (c.f., Cowan, 2001), considerable individual differences are reported in the literature and these differences are often predictive of performance on a wide variety of tasks (cf. Engle, 2001). For example, individual differences in WM capacity predict performance on attentional control (AC) tasks such as the Stroop task (Kane & Engle, 2003), the flanker task (Heitz & Engle, 2007), the antisaccade task (e.g., Unsworth et al., 2004), and dichotic listening tasks (Conway et al., 2001). Additionally, individual differences in WM capacity predict participants' ability to select task-relevant stimuli (Vogel, McCollough & Machizawa, 2005), avoid attentional capture from irrelevant stimuli (Fukuda & Vogel, 2009), and recover from failures to ignore irrelevant stimuli (Fukuda & Vogel, 2011). Research also suggests that individuals with low WM capacity show a higher proclivity toward mind wandering (Kane, Brown, McVay, Silvia, Myin-Germeys & Kwapil, 2007a) and that WM capacity predicts performance on higher-level reasoning tasks such as tests of Gf (Kane, Conway, Miura & Colflesh, 2007b) as well as Scholastic Aptitude Test (SAT) performance (Turner & Engle, 1989). This list documents only a small sample of the vast variety of tasks for which individual differences in WM capacity can predict performance. Thus, it is no surprise that researchers are interested in developing methods to improve WM capacity which may then lead to improvements in other cognitive skills.

Cognitive training, therefore, should have two goals (Willis, Tennstedt, Marsiske, Ball, Elias & Koepke, 2006): to improve performance on the training task itself; and to improve performance on untrained tasks and everyday functioning. While the cognitive training literature unequivocally reports improvement on the trained tasks, improvement on untrained transfer tasks is less reliable and intermittently reported in the literature (see Morrison & Chein, 2011, for a review).

While the WM training literature is quickly growing, an exhaustive review is impractical in the current text (for a recent review see Melby-Lervåg & Hulme, 2013; Morrison & Chein, 2011; Shipstead et al., 2010). Briefly, however, the influence of WM training on several cognitive processes has been extensively investigated. For example, in their now seminal study, Jaeggi et al. (2008) demonstrated improvement on measures of Gf following dual *n*-back training compared to a control group that did not receive practice [i.e., no-contact control (NCC) group]. While training-related Gf improvements has been replicated (e.g., Colom, Quiroga, Shih, Martinez, Burgaleta & Martinez-Molina, 2010; Jaeggi et al., 2010b; Klingberg, Forssberg & Westerberg, 2002; Olesen, Westerberg & Klingberg, 2004; Westerberg & Klingberg, 2007), several additional studies have failed to find such improvements (e.g., Chein & Morrison, 2010; Dahlin, Neely, Larsson, Backman & Nyberg, 2008a; Owen et al., 2010; Redick et al., 2012). Similarly, several studies have reported training-related improvements on complex span measures of WM (e.g., Chein & Morrison, 2010; Dahlin et al., 2008a; Dahlin, Nyberg, Backman & Neely, 2008b; Li, Schmiedek, Huxhold, Rocke, Smith & Lindenberger, 2008; Lilienthal, Tamez, Shelton, Myerson, & Hale, 2013), while others (and in some cases those same studies) report no significant improvement (e.g., Jaeggi et al., 2008; Li et al., 2008; Lilienthal et al., 2013; Owens, Koster & Derakshan, 2013; Redick et al., 2012; Schmiedek, Lövdén & Lindenberger, 2010). The influence of training on AC is less controversial with most studies reporting successful transfer (e.g., Chein & Morrison, 2010; Klingberg et al., 2002; Olesen et al., 2004; Westerberg & Klingberg, 2007) with only two exceptions (Owen et al., 2010; Dahlin et al. 2008a). For consistency with the literature, and to help weigh in on the more controversial reported findings, the current study includes measures of Gf, WM, and AC.

While reports of the efficacy of WM training to transfer tasks of Gf, WM, and AC are widespread, the main goal of the present experiments is to specifically address WM training transfer efficacy to visual short-term memory (VSTM) performance; an under-investigated area in the brain training literature. In this study, visual short-term memory was assessed using the change detection task (e.g., Awh, Barton & Vogel, 2007) and the short-term recall task (e.g., Zhang & Luck, 2008), standard cognitive measures of VSTM function infrequently included in cognitive training batteries. The benefit of enhanced VSTM has many applications and could be particularly advantageous for air traffic controllers, system managers, machine operators, warfighters, or any employment requiring the detection of visual patterns and monitoring changes in visual displays. Finally, while extensive practice on various VSTM tasks can lead to performance improvement (Gaspar, Neider, Simons, McCarley & Kramer, 2013; Olesen et al., 2004; Zimmer, Popp, Reith & Krick, 2012), there have been very few investigations the benefit of cognitive training transfer on VSTM.

To the best of our knowledge there have been three studies to date investigating the effects of multi-session cognitive training (namely WM and AC training) on

VSTM transfer tasks, and these three studies report disparate conclusions (Owens et al., 2013; Arend & Zimmer, 2012; Kundu, Sutterer, Emrich & Postle, 2013). Arend and Zimmer (2012) trained participants on a multiple object tracking task for 10 sessions. While participants improved across training sessions, no training-related improvements were observed on a change detection transfer task suggesting that AC training does not improve VSTM performance. Two studies, however, that do report significant improvements on VSTM tasks following WM training. In an EEG experiment, Owens et al. (2013) trained dysphoric patients on either an adaptive dual *n*-back task (Jaeggi et al., 2008) or a dual 1-back task for eight sessions. VSTM performance was assessed before and after training on a change detection task (Vogel et al., 2005). Behavioral and EEG data reveal both training specific gain in VSTM capacity and improved filtering efficiency of irrelevant information following training. This study provides evidence of successful transfer of WM training to a VSTM task in a patient population. Kundu et al. (2013) show similar results in a non-patient population. Following approximately 24 days of dual *n*-back training, VSTM capacity increased compared to a control group who played Tetris for 24 days. Additionally, EEG data show a reduction in VSTM-related neural activity following training suggesting more efficient information processing.

Thus, while both Owens et al. (2013) and Kundu et al. (2013) report general increased VSTM capacity following WM training, it has been suggested that VSTM may not be a unitary construct and investigating the contribution of WM training to unique subcomponents may help explain why some studies succeed and others fail to show training-related improvement. Awh et al. (2007) have proposed a two-factor model of VSTM capacity suggesting that VSTM capacity depends critically on two independent subprocesses. According to this model, VSTM relies both on the number of items held in memory and the resolution or discriminability with which those items are stored. Additionally, VSTM number and resolution are separable processes (e.g., Awh et al., 2007; Fukuda, Vogel, Mayr & Awh, 2010; Scolari, Vogel & Awh, 2008) both of which contribute to overall VSTM capacity. As independent and separable processes, it is possible that the number and resolution would be differentially influenced by a cognitive training regimen. Furthermore, data suggest that these processes are relatively stable and neither number nor resolution measures can be enhanced with motivation, either instructional or monetary (Zhang & Luck, 2011). The effect of cognitive training on number and resolution, however, has not been evaluated. By dissociating these two subprocesses, we can determine whether training has an effect on improving the number of items held in memory, the discriminability between those items held in memory, or both.

In the experiments presented here, variants of the single *n*-back task were used during training (for a comparison with dual *n*-back training see Jaeggi et al., 2010b). The *n*-back task is a complicated working memory task that engages multiple component processes. In the standard version of the task, participants must monitor a continuous string of stimuli and respond when a given stimulus matches the stimulus that appeared *n* trials previously. In depth analyses of the requisite component processes revealed that successful *n*-back task performance requires monitoring, maintaining, and updating of memory representations, selection and interference resolution among multiple stimuli, as well as inhibitory control of irrelevant information/stimuli (Jaeggi et al., 2010a; Jonides, Schumacher, Smith, Lauber, Awh, Minoshima & Koeppe, 1997). As such, using the *n*-back task as a cognitive training tool constitutes “core training” targeting domain-general cognitive processes shared by many other tasks (Morrison & Chein, 2011). Successful VSTM performance relies on many of the same cognitive processes. Participants must monitor and maintain information across a delay, and they must inhibit remembered displays from previous trials as well as irrelevant stimuli on the current trial; in fact, filtering irrelevant information is essential for successful VSTM task performance (Owens et al., 2013; Vogel et al., 2005). Because there is considerable overlap between *n*-back and VSTM task component properties, improvement following *n*-back training that transfers to VSTM performance would be considered near transfer (Barnett & Ceci, 2002; Shipstead et al., 2010).

Experiment 1

Experiment 1 investigated the influence of WM training on the measures of VSTM number and VSTM resolution independently. Participants were trained on both a verbal and spatial version of an adaptive single *n*-back task; single *n*-back tasks have previously proven an effective cognitive training tool (e.g., Jaeggi et al., 2010b). Training performance was evaluated across eight 1-h sessions over approximately 2–4 weeks for a total of 5,760 training trials. While the number of training trials varies greatly across experiments, successful transfer has previously been shown with as few as 2,625 training trials (Jaeggi, Buschkuehl, Jonides & Shah, 2011). VSTM was tested before and after training using a particular variant of the change detection task designed to dissociate number and resolution subprocesses (Awh et al., 2007). For consistency with the literature, participants were also assessed on a pre- and post-training battery including measures of WM, Gf, and attentional control (AC).

Method

Participants

Fifty-three participants (ages 18–30; 23 women) with normal to corrected-to-normal vision from the Georgia Institute of Technology completed this experiment. All participants were paid \$10/h as compensation with a monetary bonus (up to \$10) based on training performance.

Cognitive assessments

Each participant completed a battery of computerized tasks both on the first and last sessions of the experiment. All tests were presented using a Dell Dimensions PC computer and 24″ CRT monitor using Eprime (Schneider, Eschman & Zuccolotto, 2002), Presentation® software (Version 16.1, www.neurobs.com), and NTI Armory software (O'Donnell, Moise & Schmidt, 2005). This battery was designed to assess participants' spatial and verbal WM, VSTM, Gf, and AC abilities.

Visual short-term memory measures: a change detection task was used to measure VSTM. On each trial, a centrally presented arrow (100 ms) cued participants to one side of the display. After a short delay (100 ms), a memory set was presented containing 2, 4, 6, 8, or 10 items distributed evenly across the left and right of the display (100 ms). Participants needed to direct their attention to the items on the cued side of the display and ignore or inhibit those on the uncued side. After a brief delay (900 ms), two probes were presented and remained on the screen until the participant made a response. The participants' task was to decide whether the probe on the cued side of the screen was the same or different from the item that appeared at that same location in the memory set. The stimulus set included two different vertically oriented rectangles and two different horizontally oriented ovals (described in detail by Fukuda et al., 2010) randomly selected on each trial. VSTM number was measured using accuracy on between-category trials (e.g., change from a rectangle to an oval) and VSTM resolution was assessed using accuracy on within-category (e.g., change from one oval to another oval) as described by Awh et al. (2007). A change occurred on 50 % of the trials. There were a total of 60 trials.

Working memory measures: two automated WM tasks were included in the battery, operation span and symmetry span. The automated operation span task is a complex span tasked designed to evaluate verbal WM capacity. On each trial, participants were asked to remember a string of 3–7 letters presented one at a time (1,000 ms) with intervening math problems. Participants solved the math problem and remembered the letters. At the end of the trial, participants reported the to-be-remembered letters in order. This task has been previously described in detail (Unsworth, Heitz, Schrock & Engle, 2005). There were a total of 15 trials (3 trials per string length; max score = 75).

The automated symmetry span task is a complex span task designed to measure spatial WM capacity. This task is conceptually identical to the automated operation span task except that participants must remember 2–5 spatial locations (4×4 matrix with one cell filled in red; 650 ms) and make intervening symmetry judgments (i.e., Is the presented geometric feature symmetric about vertical axis?). There were a total of 12 trials (3 trials at each string length; max score = 42).

General fluid intelligence measure: the Raven's advanced progressive matrices (RAPM) task is a measure of Gf. On each trial, participants were presented with a test set composed of 8 figures arranged in a 3×3 matrix with the bottom right cell missing. Participants were asked to determine which of the eight possible figures best fit into that missing cell. Participants were given 10 min to complete up to 18 problems (odd trials vs. even trials counterbalanced across participants). This version of RAPM has previously been used by Jaeggi et al. (2008) and Redick et al. (2012).

Attentional control measures: two AC tasks from the NTI Armory battery of cognitive tests were used to measure focused attention abilities. The NTI Armory was developed to test cognitive skills involved in real-world military missions and are used regularly in the training of U.S. Airmen (O'Donnell et al., 2005). In the motion interference task, on each trial, a white arc with a large tick mark was presented on a black background. A circle appeared and began to move along the arc at a constant trajectory before disappearing. Four letters then appeared below the arc and participants had to determine whether or not a vowel was present. Next participants were asked to stop the now-invisible-circle as close to the tick mark as possible based on its previous trajectory. Deviation scores (distance between the tick mark and where the circle was stopped) were acquired. There were a total of 60 trials.

In the rapid decision-making task, on each trial, participants were presented with three concentric circles each representing a different level of threat (i.e., center circle was the most threatening and outermost circle the least threatening). Also on each trial, three shapes (X, O, and ?) were presented inside the circles. These shapes were also assigned a level of threat (i.e., X was the most threatening and O the least threatening). The task was to determine which shape presented the greatest threat as quickly as possible depending on both the identity of the shape and its location within the circles. Reaction times (RT) were collected. There were a total of 60 trials.

Training tasks

Half of the participants completed eight sessions of adaptive *n*-back training (Fig. 1). During each session, participants completed 18 blocks of a verbal and 18 blocks of a spatial version of this task. For both versions of the task, on each trial participants were presented with a 5×5 matrix outlined in white on a black background. A capital letter (consonants only) then appeared inside one of the cells (500 ms). The participant then pushed a button (1 or 2 key on the keyboard) to indicate whether the current stimulus matched the stimulus that appeared *n* trials previously (2,500 ms). In the verbal version of the task, participants indicated whether the identity of the letter was identical to the identity of the letter that appeared *n* trials previously and ignored information about the spatial location of those letters. In the spatial version of the task, participants indicated whether the location of the current letter matched the location of the letter that appeared *n* trials previously and ignored information about the identity of that letter. The task was adaptive on a block-by-block basis. If a participant achieved 95 % or better on a given block, *n* was increased by one on the next block. Alternatively if participants achieved less than 75 % on a given block, *n* was reduced by one. Otherwise *n* remained the same. Each block consisted of $20 + n$ trials and only the last 20 trials were included when calculating block accuracy. Each session began at $n = 1$.

Results

Training task

Training task improvement was evaluated with a Task (verbal vs. spatial) × Session (1–8) repeated measures analysis of variance (ANOVA; Fig. 2) on the mean *n*-back achieved for each session. Index of effect size is reported using partial eta-squared. The assumption of sphericity was violated for both the main effect of Session ($p < 0.001$) and the Interaction ($p = 0.025$), thus degrees of freedom were corrected using the Huynh–Feldt adjustment. The main effect of Session was significant, $F(4.6,120.2) = 28.53$, $p < 0.001$, $\eta_p^2 = 0.52$, with performance improving across the 8 training sessions (Fig. 2). Neither the main effect of Task, $F(1,26) = 1.64$, $p = 0.212$, $\eta_p^2 = 0.06$, nor the Interaction, $F(6.7,174.8) = 0.262$, $p = 0.964$, $\eta_p^2 = 0.01$, was significant indicating that improvement was similar for both the verbal and spatial versions of the adaptive *n*-back task.

Cognitive assessments

Performance improvements on the cognitive battery tasks were assessed with a Time [battery session 1(BS1) vs. battery session 2(BS2)] × Group (Training vs. Control) repeated measures ANVOA on the mean performance on each task. Significant Time × Group interactions (just "Interaction" below) indicate training-related performance improvement. The dependent measures varied among the various tasks and are indicated below. One primary ANOVA was conducted to evaluate whether training improved performance on the change detection task and this finding was then decomposed into identical ANOVAs for both number and resolution subprocesses. A secondary family of ANOVAs was conducted to evaluate the effect of training on each WM, Gf, and AC tasks and these comparisons were not Bonferroni corrected. The necessity of multiple comparison correction in the cognitive training literature is contentious. Although multiple comparison corrections have been largely ignored in the literature (Chein & Morrison, 2010; Dahlin et al., 2008a; Jaeggi et al., 2008; Li et al., 2008), this practice has been criticized

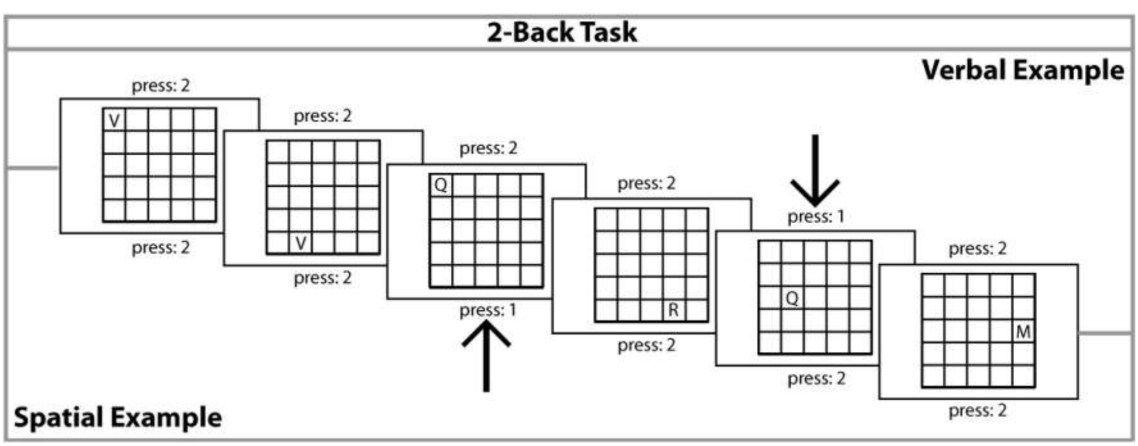

Fig. 1 Verbal (*top*) and spatial (*bottom*) versions of the single *n*-back task used during training. This is an example of a 2-back condition. *Arrows* indicate where a target occurred thus necessitating a 1-key push response

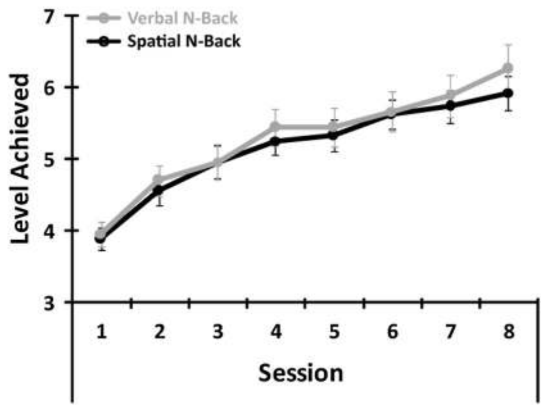

Fig. 2 Averaged group performance on the verbal and spatial versions of the *n*-back task. Participants showed improvement across the 8 training sessions. *Standard error bars* are shown

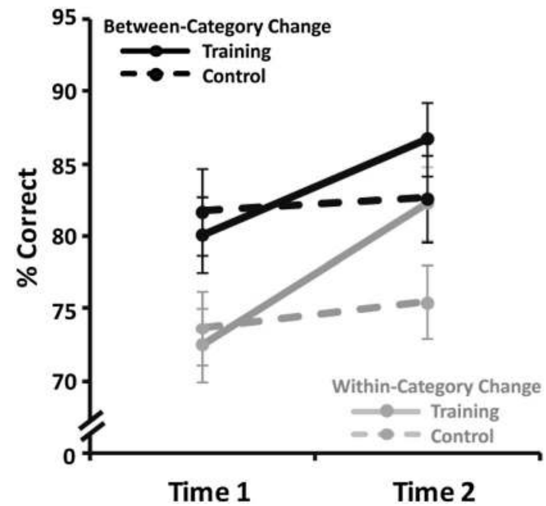

Fig. 3 Training-related performance improvement on the change detection task separated by type of change that occurred. Between-category change trials indicate performance improvements in the number of items held in visual working memory and within-category change trials indicate performance improvements in the resolution with which those items are stored. *Percent correct* and *standard error bars* are shown

and studies are beginning to use statistical correction as a means to protect against type I error (e.g., Lilienthal et al., 2013; Redick et al., 2012). However, with one primary analysis and a family of secondary analyses and an experiment designed specifically to test these comparisons (as in the current study), it has been suggested that multiple comparison correction may not be appropriate or necessary (Motulsky, 2010). We, therefore, do not report Bonferroni-corrected significance values here (however, for the interested reader, none of the significant effects survived such correction in Experiment 1). Index of effect size is reported using partial eta-squared.

Primary battery task

Visual short-term memory task: for the change detection task, the dependent variable of interest was accuracy. Two participants were removed from the analysis (one from each group due to technical failures during the final battery session). The main effect of Time, $F(1{,}49) = 8.49$, $p = 0.005$, $\eta_p^2 = 0.15$, was significant. The main effect of Group, $F(1{,}49) = 0.40$, $p = 0.529$, $\eta_p^2 = 0.01$, was not significant. The Interaction, $F(1{,}49) = 4.28$, $p = 0.044$, $\eta_p^2 = 0.08$, was significant. This interaction suggests that *n*-back training improved change detection performance (replicating Kundu et al., 2013; Owens et al., 2013).

Improved change detection following training was further evaluated by separating out the between-category (i.e., number) and within-category (i.e., resolution) components consistent with the two-factor model of VSTM (Awh et al., 2007). The influence of *n*-back training on each of these processes was assessed separately. The between-category ANOVA (Fig. 3) revealed that neither the main effect of Group, $F(1{,}49) = 0.18$, $p = 0.677$, $\eta_p^2 = 0.004$, nor the Interaction, $F(1{,}49) = 1.93$, $p = 0.172$, $\eta_p^2 = 0.04$, was significant. The main effect of Time approached significance, $F(1{,}49) = 3.64$, $p = 0.062$, $\eta_p^2 = 0.07$. The within-category ANOVA revealed a significant main effect of Time, $F(1{,}49) = 10.14$, $p = 0.003$, $\eta_p^2 = 0.17$, and significant Interaction, $F(1{,}49) = 4.85$, $p = 0.032$, $\eta_p^2 = 0.09$. The main effect of Group was not significant, $F(1{,}49) = 0.58$, $p = 0.451$, $\eta_p^2 = 0.01$.

These data demonstrate that global change detection performance improved after training. When number and resolution components were considered separately, there was some evidence that the training-related improvement was larger for within-category changes (i.e., measure of resolution); however, between-category trials showed a similar trend. In fact, when the training-related improvement for within- and between-category trials was directly assessed with a Time × Group × Change Type (within- vs. between-category) ANOVA, the 3-way interaction was not significant, $F(1{,}49) = 0.371$, $p = 0.545$, $\eta_p^2 = 0.01$, suggesting similar improvement for number and resolution.

Additional battery tasks

Additional battery tasks investigating the effect of WM training on two WM tasks, two AC tasks, and one Gf task were included for consistency with the existing literature. Details are outlined below; however, these data indicate that cognitive training improvement did not transfer to of these tasks with the exception of the automated operation span task.

Working memory tasks: for the automated operation span task, the dependent variable of interest was the total score (max. 75). Four participants (two from each group) were removed from the analysis because they made greater than 15 % errors on the math task. The ANOVA revealed a significant main effect of Time, $F(1,48) = 10.52$, $p = 0.002$, $\eta_p^2 = 0.18$ and a significant Interaction, $F(1,48) = 6.28$, $p = 0.016$, $\eta_p^2 = 0.12$, The main effect of Group, $F(1,48) = 2.47$, $p = 0.122$, $\eta_p^2 = 0.05$, was significant. For the automated symmetry span task, the dependent variable of interest was again the total score (max. 42). Two participants (one from each group) were removed from the analysis because they made greater than 15 % errors on the symmetry judgment task. The ANOVA again revealed a significant main effect of Time, $F(1,49) = 17.08$, $p < 0.001$, $\eta_p^2 = 0.26$, but neither the main effect of Group, $F(1,49) = 2.16$, $p = 0.148$, $\eta_p^2 = 0.04$, nor the Interaction, $F(1,49) = 0.63$, $p = 0.432$, $\eta_p^2 = 0.01$, was significant. Raw difference scores (BS1 − BS2) are listed in Table 1. In this experiment, *n*-back training did not improve WM.

Fluid intelligence task: for the RAPM task, the dependent variable of interest was the total score (max. 18). The ANOVA indicated that neither the main effect of Time, $F(1,51) = 1.03$, $p = 0.315$, $\eta_p^2 = 0.02$, nor the Interaction, $F(1,51) = 0.54$, $p = 0.467$, $\eta_p^2 = 0.01$, was significant. The main effect of Group, $F(1,51) = 4.50$, $p = 0.039$, $\eta_p^2 = 0.08$, was significant with the training group scoring higher than the NCC group. Raw difference scores (BS1 − BS2) are listed in Table 1. In this experiment, *n*-back training did not improve Gf.

Attentional control tasks: performance on the rapid decision-making task was measured via RT. The ANOVA revealed a significant main effect of both Time, $F(1,51) = 36.86$, $p < 0.001$, $\eta_p^2 = 0.42$, and Group, $F(1,51) = 17.48$, $p < 0.001$, $\eta_p^2 = 0.26$. The Interaction, $F(1,51) = 0.60$, $p = 0.442$, $\eta_p^2 = 0.01$, was not significant. Performance on the motion interference task was measured via deviation score (i.e., root mean square) for stopping the circle. The motion interference task proved difficult for participants, and only participants who both stopped the circle and responded to the vowel task were included in the analyses. Consequently, 5 and 10 participants were removed from the Training and Control groups, respectively. The ANOVA again revealed a significant main effect of Time, $F(1,36) = 9.13$, $p = 0.005$, $\eta_p^2 = 0.20$, but neither the main effect of Group, $F(1,36) = 0.08$, $p = 0.782$, $\eta_p^2 = 0.002$, nor the Interaction, $F(1,36) = 2.79$, $p = 0.104$, $\eta_p^2 = 0.07$, was significant. Raw difference scores (BS1 − BS2) are listed in Table 1. In this experiment, *n*-back training did not improve real-world military relevant measures of AC.

Discussion

The present data add to a growing body of literature suggesting that cognitive training may improve performance on untrained cognitive tasks; however, the benefit of training is likely process specific (Dahlin et al., 2008a, b; Melby-Lervåg & Hulme, 2013). The present data are consistent with the Owens et al. (2013) and Kundu et al. (2013) studies showing that training improves performance on a VSTM task. Our data, however, go beyond previous work to demonstrate that cognitive training may improve number and resolution subprocesses that contribute to VSTM performance. While only the training effect on VSTM resolution was significant, visual inspection of the VSTM number data hint at a trend in the same direction. The number comparison had less power (observed power = 0.28) than the resolution comparison (observed power = 0.58), and therefore further research with larger sample sizes and more power is necessary to tease apart this difference if it exists. Thus, these data do not provide definitive support for subprocess-specific training improvement. With the exception of the automated operation span task, performance on no other tasks showed significant improvement following training. Overall, these data suggest that *n*-back training may lead to very specific cognitive performance improvements and that improvements may not be as far reaching as previously suggested (e.g., Jaeggi et al., 2008, Jaeggi et al., 2010b).

Table 1 Difference scores for battery session 1 and battery session 2

	Battery session 2–battery session 1; difference score	
	Control Group	Training group
Automated operation span	1.1 (13.60)	10.9 (14.2)
Automated symmetry span	3.9 (8.3)	6.9 (10.1)
General fluid intelligence		
Raven's advanced progressive matrices	0.69 (3.2)	0.11 (2.5)
Attentional control		
Motion interference[a]	0.13 (0.15)	0.04 (0.19)
Rapid decision-making[a]	300.7 (429.7)	233.2 (156.3)

[a] Reverse scoring

There are four major limitations of this experiment. First, the comparison group in this study was a NCC group. It has been argued that a NCC group may not be the most valid comparison group due to differences in motivation, expectations, and/or demand characteristics (Shipstead et al., 2010). Second, WM training included both a verbal and spatial variant of the *n*-back task so improvement can only be attributed to this combination and not to either verbal or spatial training more specifically. Third, VSTM and Gf were all assessed using a single task. While changing scores on single tests may be interesting, the ultimate goal of brain training should be to alter general cognitive abilities (Shipstead et al., 2010; Morrison & Chein, 2011). Using multiple measures of a given process reduces the influence of any task-specific findings leading to a more robust measure of cognitive ability (Redick et al., 2012). These three limitations were addressed in Experiment 2. Finally, while we report significant transfer to VSTM in general and VSTM resolution specifically, non-significant trends in the data suggest the necessity for a more powerful training design. Therefore, specific modifications were made to the training task itself to promote better learning across training sessions and perhaps more improvement on the transfer tasks.

Experiment 2

Given recent enthusiasm surrounding cognitive training and the spike in the number of publications on the topic (viz. Morrison & Chein, 2011; Melby-Lervåg & Hulme, 2013; Redick et al., 2012), it is easy to forget that this is not, in fact, a new area of research. Psychologists have been interested in understanding learning and training for over a century (Ebbinghaus, 1885/1913). Specific recommendations regarding feedback, duration, and organization for optimized training designs have been outlined in detail in the skill learning literature (e.g., Adams, 1987; Bartlett, 1947; Dempster, 1988; Kerr & Booth, 1978; Pashler, Rohrer, Cepeda & Carpenter, 2007; Rogers, 1996; Schmidt & Bjork, 1992; Schneider & Chen, 2003; Shea & Morgan, 1979).

One recommendation from this literature is that the adaptive training is essential for the success of cognitive training regimens (e.g., Jaeggi et al., 2008; Jaeggi et al., 2010b; Klingberg et al., 2002; Lilienthal et al., 2013; Brehmer, Westerberg & Bäckman, 2012; Chein & Morrison, 2010; Schneiders, Opitz, Krick & Mecklinger, 2011, Schneiders, Opitz, Tang, Deng, Xie & Li, 2012). In these adaptive designs, task difficulty is determined by task performance. When participants perform well in a given block, task difficulty increases on the next block. When participants perform poorly, the difficulty of the subsequent block is reduced. However, when participants perform moderately well (typically between 75 and 95 % accurate), task difficulty remains constant. This performance-based difficulty adjustment is designed to create variability on storage demands in the training environment which may be essential for successful skill transfer (cf. Schmidt & Bjork, 1992). Experiment 1 used this design; however, we noticed some unexpected and interesting trends in these data. In our data set, it was very common for participants to plateau at a certain level of difficulty and remain at that level for several consecutive blocks. Thus, when participants consistently performed moderately well, task difficulty remained stagnant and storage-demand variability substantially decreased. In fact, on 59.1 % of training sessions, task difficulty remained constant for greater than five consecutive blocks (more than one quarter of the session) and on 21.9 % of training sessions, task difficulty remained constant for half of the session (9+ consecutive blocks). Thus the essential task variability may, in fact, be minimized in the standard adaptive design, at least in this experiment. As aggregate data across session are typically reported in the cognitive training literature, it is impossible to determine whether our data are unique; however, we suspect that this pattern is not unusual. Experiment 2 was, thus, designed to avoid performance plateaus and promote storage-demand variability as intended by the standard adaptive design.

Training task requirements were varied in Experiment 2 by incorporating a forced-adaptive version of the *n*-back tasks to ensure that difficulty remained variable across each training session. In this forced-adaptive design, task difficulty was obligatorily increased when participants spent more than five consecutive blocks at the same level of difficulty. If participants were less than 75 % accurate on this block, the data were not included when calculating highest level of performance and task difficulty returned to the previous level on the subsequent block. Such forced variability should enhance retention (Schmidt & Bjork, 1992) and encourage transfer (Schneider et al., 2002).

In Experiment 2, we further investigated the effect of cognitive training on VSTM generally and its component parts (i.e., number and resolution) using multiple VSTM tasks. Additionally, this experiment sought to ameliorate the limitations of Experiment 1 by (1) including multiple measures for each WM and Gf, and including more traditional psychological measures of AC, (2) by including a contact control comparison group and larger sample size, and (3) by using a forced-adaptive design to optimize training.

Method

Participants

Sixty-nine naïve volunteers (ages 18–32; 31 women) were recruited from the Georgia Institute of Technology community for this experiment. For their participation, participants received either pay ($10/h) or course credit (1 credit/h) in partial fulfillment of a course requirement with a monetary bonus (up to $10) based on training performance.

Groups

Participants were randomly assigned to three groups: a NCC group, a verbal training (VT) group, and a spatial training (ST) group. All participants completed a test battery session on the first and last days of participation. Battery tasks are outlined below. In addition, both the VT and ST groups completed the eight intervening training sessions. Training sessions included 40–60 min of the adaptive verbal *n*-back and adaptive spatial *n*-back task, respectively (described in detail below).

Cognitive assessments

As in Experiment 1, a battery of computerized tasks was administered during the first and the last experimental sessions. All tests were presented using a Dell Dimensions PC computer and 24″ CRT monitor using Eprime (Schneider et al., 2002) or MATLAB. The battery tasks included two tests of VSTM (change detection and short-term recall tasks), two tests of WM (automated operation span and automated symmetry span tasks), two tests of Gf [i.e., RAPM and Cattell's Culture Fair (CCF) tasks], and two tests of AC (flanker and antisaccade tasks).

Visual short-term memory measures: the Change Detection task was modified from that used in Experiment 1. In Experiment 1, 30 % of the participants were more than 90 % accurate during BS1. The Change Detection task was, therefore, modified in Experiment 2 to increase the difficulty and allow greater opportunity for post-training improvement (timing based on Awh et al., 2007; stimuli from Fukuda et al., 2010). In this experiment, on each trial, a centrally presented arrow (200 ms) cued participants to one side of the display. After a short delay (200 ms), a memory set was presented containing 2, 4, 6, 8, 12, or 16 items distributed evenly across the left and right of the display (500 ms). Again, participants had to focus on those shapes presented on the cued side of the screen and ignore those on the uncued side. After a brief delay (1,000 ms), two probes were presented (test display: 2,000 ms). The participants' task was to decide whether the probe on the cued side of the screen was the same or different from the item that appeared at that same location in the memory set and indicate their decision with a button push. The stimulus set was identical to that used in Experiment 1 and Fukuda et al. (2010). Again between-category and within-category trials were used to differentiate between VSTM number and resolution (Awh et al., 2007; Fukuda et al., 2010). Participants completed 20 practice trials (5 within category, 5 between category, and 10 no change) followed by three experimental blocks each with 48 trials (12 within category, 12 between category, and 24 no change). At the end of each block, participants were shown both their mean speed and accuracy on that block. This is a measure of VSTM capacity.

On each trial of the short-term recall task (Zhang & Luck, 2008) six capital letters (A, B, C, D, E, F, or G) appeared at random locations around a fixation point (100 ms). After a delay (900 ms), a gray wheel (same dimensions as the color wheel from the color version of the experiment) appeared and one of the previously presented letters appeared in the center. Participants used the mouse to click on the gray wheel indicating at which location that letter had previously appeared. Again, participants completed five practice trials followed by four experimental blocks with 60 trials each. This too is a measure of VSTM capacity.

Working memory measures: automated operation span and automated symmetry span tasks were identical to Experiment 1.

General fluid intelligence measures: RAPM task was identical to Experiment 1.

In the CCF task, participants completed four subtasks (series completion, odd elements, matrix completion, and dot task; Cattell, 1949). In the series completion task (7 problems per session), participants saw three simple line drawings that together created a pattern. Participants had to decide which of six possible similar pictures best completes the pattern. In the odd elements task (7 problems per session), participants saw five simple line drawings and had to determine which two drawings did not belong with the rest. In the matrix completion task (7 problems per session), participants were presented with either a 2 × 2 or 3 × 3 matrix. One of the cells was empty and participants had to decide which of four possible alternatives best fit into the empty cell. Occasionally the matrices were partially obscured by "cut outs" or missing information. In the dot task (6 problems per session), participants were presented with a simple line drawing with a dot present. Participants had to then determine in which of the five possible alternative drawings (without dots) would allow for a dot to be placed in a comparable position to the sample drawing. Participants completed odd numbered problems during one session and even numbered problems during the other session (order was counterbalanced across participants) for a total of 27 problems per session. This is also a measure of Gf.

Attentional control measures: in the antisaccade task (Unsworth et al., 2004), each trial began with a blank screen (400 ms) followed by a three asterisk fixation (200, 600, 1,000, 1,400, and 1,800 ms) and another blank screen (10 ms). Next a cue appeared (equal sign; two blinks with 100 ms on and 50 ms off) either on the left or the right of the display. A target letter (B, P, or R) appeared briefly (100 ms) on the opposite side of the display and was immediately masked with an H (50 ms) followed by an 8, which remained on the screen until the participant responded. Participants completed 60 practice trials in which cues and targets were all centrally presented with feedback on each trial. Participants also practiced 10 trials of the experimental task with feedback on each trial. Finally, participants completed 60 experimental trials with block feedback at the end; 30 with a left cue and 30 with a right cue. The three target letters as well as the five fixation durations were evenly distributed across trials. This is a measure of AC.

In the flanker task (Eriksen & Eriksen, 1974), participants were presented with a centrally presented fixation dot (200 ms) followed by five arrows (e.g., >>>>>; 100 ms). Participants were asked to determine whether the central arrow was facing the right or left and respond with a button push. Half of the trials were congruent (e.g., >>>>>) and half of the trials were incongruent (e.g., gt;><>>). There were an equal number of left and right responses. Five delays separating each trial were evenly distributed across trials (200, 600, 1,000, 1,400, and 1,800 ms). There were a total of 16 practice trials with feedback on each trial and 120 experimental trials with block feedback at the end. This is also a measure of AC.

Training tasks

The VT group completed eight sessions of the forced-adaptive verbal *n*-back task (Fig. 4). This is a continuous performance tasks in which participants must monitor a string of centrally presented letters. Throughout the experiment the outline of a white square was centrally presented on a black background. On each trial, a capital letter printed in white appeared inside the square. After 500 ms, the letter disappeared and participants had 2,500 ms to make a response before the next trial began. On each trial, participants were asked to decide if the letter presented on the current trial matched the letter that appeared *n* trials ago. This task was adaptive in that the difficulty level changed based on participant performance. If a participant achieved 95 % or better on a given block, the level of *n* was increased by one on the next block. Alternatively if participants achieved less than 75 % on a given block, the level of *n* was reduced by one. Otherwise, the level of *n* remained the same. Also, given the importance of variability in the training environment (cf. Schmidt & Bjork, 1992), if the participant completed five blocks in a row at the same level of *n,* then *n* increased by one on the following block. This forced-adaptive block was only included in the analysis if participants were able to perform within the 75–95 % accuracy range. Each block was comprised of 20 + *n* trials and only the last 20 trials were scored (because the first *n* trials were necessarily "no match" trials). There were 30 blocks per training session for a total of 4,800 scored trials across 8 days of training. Accuracy and RT feedback was provided at the end of each block and participants were verbally encouraged by the experimenter every 4–5 blocks.

The ST group completed eight sessions of the forced-adaptive spatial *n*-back task (Fig. 4). This task was

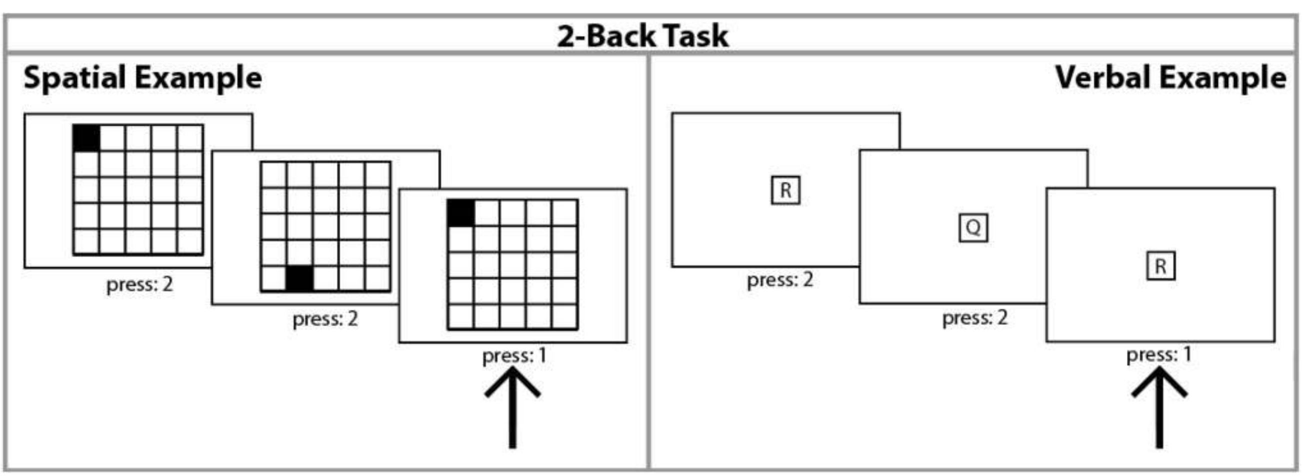

Fig. 4 Spatial (*left*) and verbal (*right*) versions of the single *n*-back task used during training. This is an example of a 2-back condition. *Arrows* indicate where a target occurred thus necessitating a 1-key push response

conceptually identical to the forced-adaptive verbal *n*-back task, except that the stimuli were spatial locations instead of letters. On each trial, a 5 × 5 grid (white on a black background) was presented and one of the cells were filled in red. After 500 ms, the filled cell disappeared and participants had 2,500 ms to make a response before the next trial began. On each trial, participants were asked to decide if the spatial location indicated on the current trial matches the spatial location that appeared *n* trials ago. Again, this task was adaptive in that the difficulty level changed based on participant performance. The adaptive schedule was identical to that used in the forced-adaptive verbal *n*-back task. Each block was comprised of 20 + *n* trials, and again only the last 20 trials were scored There were 30 blocks per training session for a total of 4,800 scored trials across 8 days of training. Accuracy and RT feedback was provided at the end of each block and participants were verbally encouraged by the experimenter every 4–5 blocks.

For both training groups, during the first training session, participants were given extensive task instructions and completed three practice blocks (1-back, 2-back, and 3-back) which were identical to the experimental blocks except that a tone sounded when an error was made. As in Experiment 1, each training session began at $n = 1$.

Procedure

During the first session, each participant completed all battery tasks over the course of approximately 2.5 h. The order of tasks was randomized across participants. Participants in the two training groups came back to the laboratory for eight additional training sessions (40–60 min each). All participants completed their final session 14–33 days after their first session (matched groups). This final session was identical to the first session except that the order of battery tasks was again randomized across participants.

Results

Two participants (one from the NCC group and one from the ST group) failed to complete the all of the required sessions and were removed from the analysis. One additional participant from the NCC group was removed from the analysis because she failed to comply with instructions on three (automated operation span, automated symmetry span, and short-term recall) of the eight battery tasks and performed greater than two standard deviations below the mean during BS1 on four (flanker, CCF, RAPM, and change detection) of the remaining five battery tasks.

Training tasks

As in Experiment 1, the effect of training on the training task itself was evaluated with a Training Session (1–8) × Group (VT, ST, NCC) repeated measures ANOVA on the maximum difficulty achieved (i.e., max *n*) on each training day (Fig. 5a). The assumption of sphericity was violated for the main effect of Session ($p < 0.001$), thus degrees of freedom were corrected using the Huynh–Feldt adjustment. The main effect of Session was significant, $F(3.8,162.1) = 62.33$, $p < 0.001$, $\eta_p^2 = 0.60$, with performance improving across the 8 training sessions. The main effect of Group was also significant, $F(1,42) = 6.01$, $p = 0.018$, $\eta_p^2 = 0.13$, with the VT group achieving an overall higher level of *n* than the ST group. Finally, the Interaction approached significance, $F(3.9,162.1) = 2.22$, $p = 0.072$, $\eta_p^2 = 0.05$, with a trend toward the VT group showing larger improvement over the course of training compared to the ST group.

When only the overall gain across training was considered (i.e., maximum level of *n* on training session 8 minus maximum level of *n* on training session 1), a two-tailed independent samples *t* test revealed a significant difference between the groups, $t(42) = 2.56$, $p = 0.015$ (Fig. 5b). These data suggest that overall, the VT group showed greater improvement from training session 1 to training session 8 (max *n* increased from 4.8 to 9.5) compared to the ST group (max *n* increased from 4.5 to 7.8), though both groups showed a benefit of training.

Cognitive assessments

To assess the efficacy of cognitive training on untrained measures of VSTM, WM, AC, and Gf, performance scores extracted separately for BS1 and BS2. Scores were then submitted to multiple Time (BS1 vs. BS2) × Group (NCC vs. ST vs. VT) repeated measures ANOVAs; one fore each task. As in Experiment 1, the primary analyses were designed to evaluate VSTM performance and a family of secondary analyses was performed to assess training-related improvement on WM, AC, and Gf task performance. As in Experiment 1, interaction results were not corrected for multiple comparisons given that primary and secondary families of ANOVAs were planned (Motulsky, 2010),[1] Index of effect size is reported using partial eta-squared.

As in Experiment 1, to further investigate the underlying processes of VSTM, number and resolution measures were assessed separately. To assess VSTM number, between-

[1] Furthermore, to preview the results of Experiment 2, if our conclusions are correct and it is improved attentional control (and more specifically, improved inhibition) driving improvement in many of our transfer measures, then these tests are not independent and Bonferroni correction is not appropriate (McDonald, 2008).

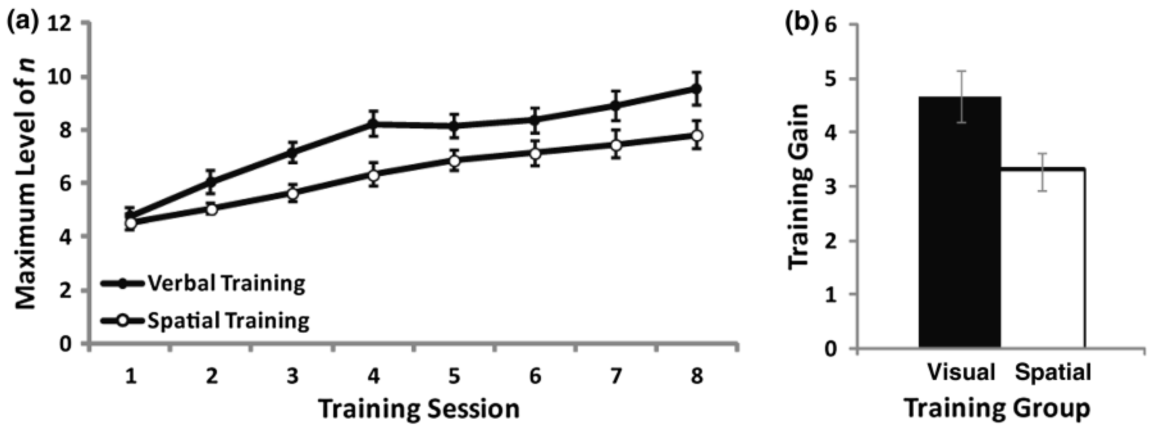

Fig. 5 **a** Averaged group performance and *standard error bars* on the adaptive *n*-back task for both the verbal and spatial training groups across training sessions. **b** Training gain from session 1 to session 8 for the verbal and spatial training groups. *Standard error bars* are shown

category accuracy scores were extracted from the change detection task (Awh et al., 2007) and Pmem (the probability that the cued item exists in memory; measure of number of items held in VSTM) measures were extracted from the short-term recall task (Zhang & Luck, 2008, 2009). To assess VSTM resolution, within-category accuracy scores were extracted from the change detection task (Awh et al., 2007) and SD (width of the von Mises distribution; measure of resolution of items held in VSTM) measures were extracted from the short-term recall task (Zhang & Luck, 2008, 2009).

Primary battery tasks

Visual short-term memory tasks: while the change detection task provides a measure of overall VSTM performance in addition to measures of number and resolution, the short-term recall task only provides independent number and resolution measures; thus, overall VSTM is only evaluated with the change detection data. In the change detection task, the dependent variable of interest was accuracy. One participant (NCC group) performed greater than three standard deviations below the mean during the first battery session and was removed from the analysis. Data were submitted to a Time x Group repeated measures ANOVA. The main effect of Group, $F(2,62) = 1.77$, $p = 0.179$, $\eta_p^2 = 0.05$, was not significant; however, both the main effect of Time, $F(1,62) = 11.01$, $p = 0.002$, $\eta_p^2 = 0.15$, and the Interaction, $F(2,62) = 6.69$, $p = 0.002$, $\eta_p^2 = 0.18$, were significant. Post hoc evaluation revealed a significant difference between the ST and NCC groups, $t(41) = 3.18$, $p = 0.003$, and the VT and NCC, $(41) = 2.56$, $p = 0.014$. The difference between the VT and ST groups, $t(42) = -0.75$, $p = 0.461$, was not significant. These data suggest that both ST and VT training improved performance on the change detection task.

Number vs. resolution: to independently evaluate the influence of brain training on VSTM number and resolution subprocesses, the change detection data were reanalyzed. Number and resolution measures were also computed from short-term recall task performance. Both the change detection and short-term recall tasks are used frequently in the literature to assess VSTM.

First, we consider measures of VSTM number. In the change detection task, training-related improvements were evaluated using a Time × Group repeated measures ANOVA on between-category change trial accuracy (Fig. 6). Both the main effect of Time, $F(1,62) = 12.25$, $p = 0.001$, $\eta_p^2 = 0.17$, and the Interaction, $F(2,62) = 10.18$, $p < 0.001$, $\eta_p^2 = 0.25$, were significant. The main effect of Group approached significance, $F(2,62) = 2.99$, $p = 0.058$, $\eta_p^2 = 0.09$. Post hoc evaluation of the efficacy of training revealed a significant difference between the VT

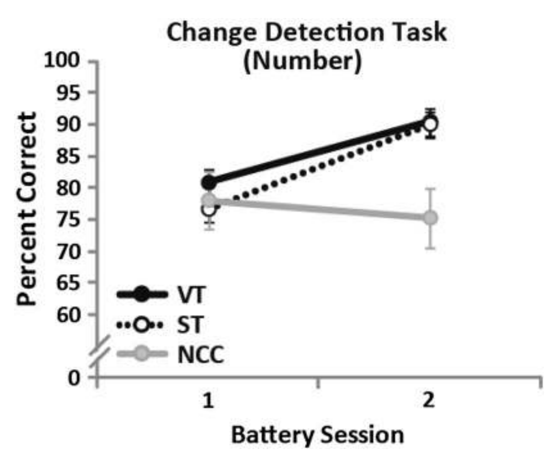

Fig. 6 Training-related performance improvement on the change detection task measure of VSTM number. Averaged percent correct for between-category trials and *standard error bars* are shown at BS1 and BS2

and NCC groups, $t(41) = 3.02$, $p = 0.002$, and the ST and NCC groups, $t(41) = 3.84$, $p < 0.001$. There was no difference between the VT and ST groups, $t(42) = -1.27$, $p = 0.210$. These data indicate that VSTM number improved following both verbal and spatial *n*-back training in the change detection task.

The specific number of items a person is capable of holding in VSTM can also be investigated by calculating capacity estimates at every set size. Thus, Cowan's K (K = [set size × (hit rate + correct rejection rate − 1)]; Cowan, 2001) was calculated at each set size for both BS1 and BS2. Separate Time (BS1 vs. BS2) × Set Size (1, 2, 3, 4, 6, and 8) repeated measures ANOVAs were conducted for each of the three groups. For the VT group (Fig. 7a), the main effect of Time, $F(1,21) = 6.42$, $p = 0.019$, $\eta_p^2 = 0.23$, and the main effect of Set Size, $F(1.9,39.1) = 11.90$, $p < 0.001$, $\eta_p^2 = 0.36$, were significant and the Interaction, $F(2.0,41.4) = 2.29$, $p = 0.115$, $\eta_p^2 = 0.1$, was not significant. For the ST group (Fig. 7b), the main effect of Time, $F(1,21) = 21.86$, $p < 0.001$, $\eta_p^2 = 0.51$, the main effect of Set Size, $F(1.9,40) = 11.20$, $p < 0.001$, $\eta_p^2 = 0.35$, and the Interaction, $F(2.9,61.2) = 4.57$, $p = 0.006$, $\eta_p^2 = 0.18$, were all significant. Finally, for the NCC group (Fig. 7c), the main effect of Set Size, $F(2.2,44.3) = 4.46$, $p = 0.014$, $\eta_p^2 = 0.182$, was significant, but neither the main effect of Time, $F(1,20) = 0.28$, $p = 0.603$, $\eta_p^2 = 0.01$, nor the Interaction, $F(2.2,43.1) = 0.38$, $p = 0.698$, $\eta_p^2 = 0.12$, was significant. Together, these data mirror the overall accuracy measures suggesting training-related improvement on VSTM number following spatial and verbal *n*-back training.

For the short-term recall task, training efficacy on VSTM number was assessed by submitting Pmem scores to a Time × Group repeated measures ANOVA (Fig. 8b). Neither the main effect of Time, $F(1,61) = 0.76$, $p = 0.388$, $\eta_p^2 = 0.01$, the main effect of Group, $F(2,61) = 2.60$, $p = 0.082$, $\eta_p^2 = 0.08$, nor the Interaction, $F(2,61) = 0.87$, $p = 0.422$, $\eta_p^2 = 0.03$, was significant suggesting no effect of *n*-back training. Similar results were obtained when capacity estimates were calculated (Zhang & Luck, 2008, 2011) and submitted by a Time x Group repeated measures ANOVA. Again, neither the main effect of Time, $F(1,61) = 0.75$, $p = 0.389$, $\eta_p^2 = 0.01$, the main effect of Group, $F(2,61) = 2.60$, $p = 0.083$, $\eta_p^2 = 0.08$, nor the Interaction, $F(2,61) = 0.87$, $p = 0.422$, $\eta_p^2 = 0.03$, was significant. These data indicate that the number of items held in VSTM as measured on the short-term recall task did not significantly change with training.

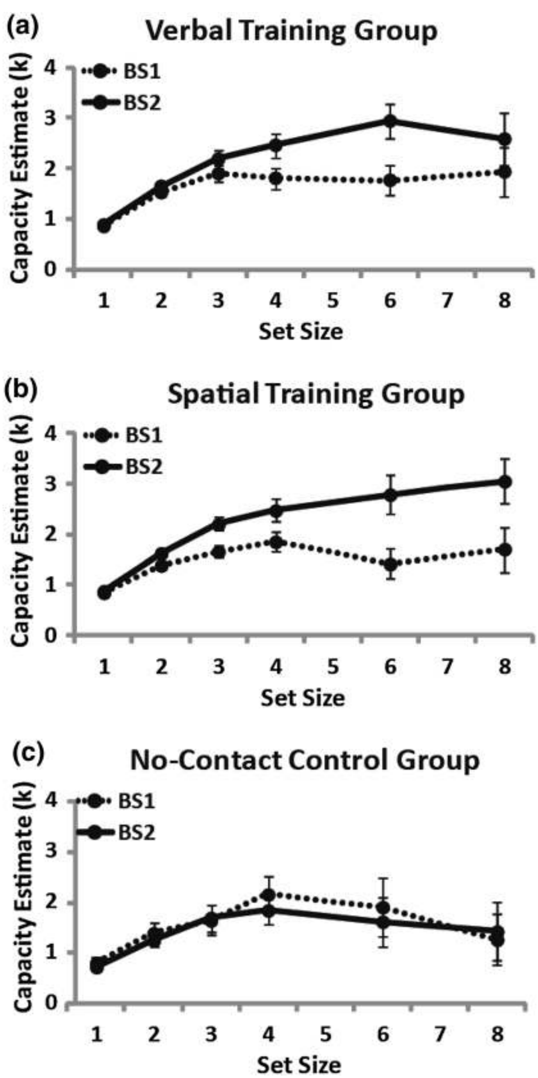

Fig. 7 Visual short-term memory capacity estimates and *standard error bars* at BS1 and BS2 for the **a** verbal training group, **b** spatial training group and **c** no-contact control group

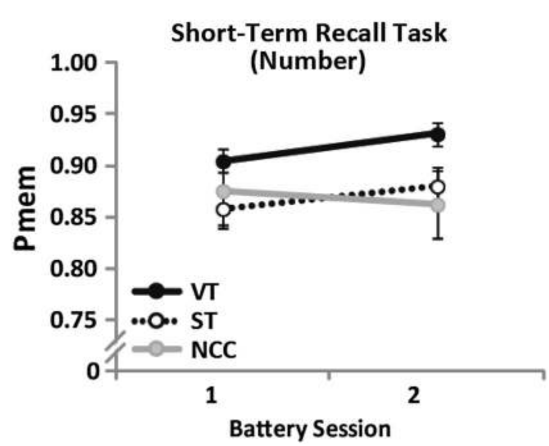

Fig. 8 Training-related performance improvement on the short-term recall task measure of VSTM number. Averaged Pmem parameters representing the probability that a probed item was present in the memory set and *standard error bars* are shown at BS1 and BS2

Next, we consider measures of VSTM resolution. In the change detection task, training-related improvements on resolution were evaluated using a Time × Group ANOVA on within-category change trial accuracy (Fig. 9a). The main effect of Time, $F(1,62) = 13.47$, $p = 0.001$, $\eta_p^2 = 0.18$, the main effect of Group, $F(2,62) = 3.70$, $p = 0.030$, $\eta_p^2 = 0.11$, and the Interaction, $F(2,62) = 7.05$, $p = 0.002$, $\eta_p^2 = 0.18$, were all significant. Again, post hoc evaluation of the efficacy of training revealed a significant difference between the VT and NCC groups, $t(41) = 2.68$, $p = 0.006$, and the ST and NCC groups, $(41) = 3.08$, $p = 0.002$, but no significant difference between the VT and ST groups, $t(41) = -0.47$, $p = 0.642$. Thus, these data indicate that both verbal and spatial *n*-back training improved performance on resolution measures in the change detection task.

In the short-term recall task, training-related improvements on resolution were tested using a Time × Group ANOVA on SD measures (note that a smaller SD measure indicates better performance; Fig. 9b). Both the main effect of Time, $F(1,61) = 4.39$, $p = 0.040$, $\eta_p^2 = 0.07$, and the Interaction, $F(2,61) = 6.31$, $p = 0.003$, $\eta_p^2 = 0.17$, were significant. The main effect of Group, $F(2,61) = 1.31$, $p = 0.277$, $\eta_p^2 = 0.04$, was not significant. Post hoc evaluation revealed a significant difference between the VT and ST groups, $t(40) = 3.20$, $p = 0.004$, as well as between the VT and NCC groups, $t(40) = 3.29$, $p = 0.001$. The difference between ST and NCC groups was not significant, $t(40) = 0.40$, $p = 0.346$. These data indicate that both the ST and NCC groups showed similar improvement from BS1 to BS2 and the VT group performed worse on BS2 compared to BS1. These data suggest that in the short-term recall task, resolution did not improved following *n*-back training.

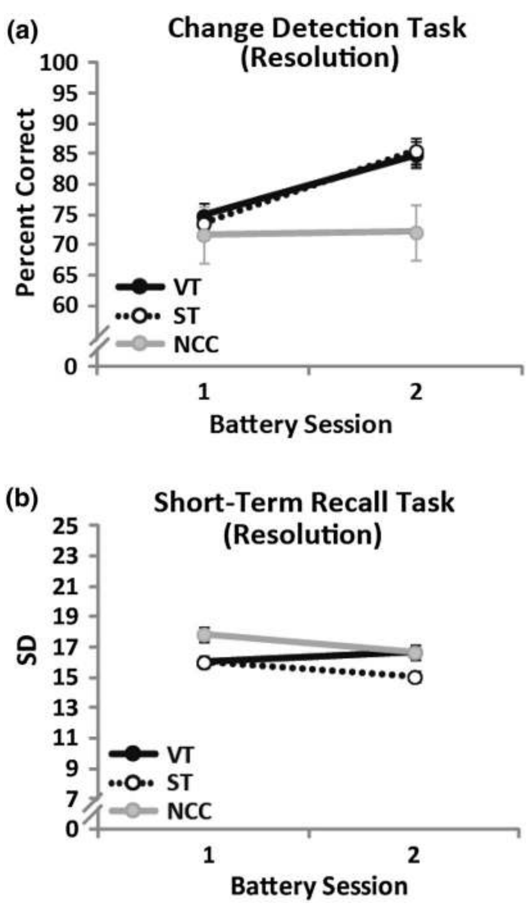

Fig. 9 Training-related performance improvement on **a** change detection and **b** short-term recall task measures of VSTM resolution. For the change detection task, averaged percent correct for within-category trials and *standard error bars* are shown at BS1 and BS2. For the short-term recall task, averaged SD parameter representing the resolution of the representations held in memory and *standard error bars* are shown at BS1 and BS2

Additional battery tasks

Additional battery tasks investigating the effect of WM training on WM, AC, and Gf were included for consistency with the existing literature. These constituted the secondary family of analyses. Training-related transfer was not evident for any of the tasks assessed. Raw difference scores (BS1-BS2) are listed in Table 2.

Working memory tasks: for both the automated operation span and automated symmetry span tasks, training-related performance improvement was measured by comparing the total score of correct items from BS1 to the total score of correct trials from BS2. Two participants were removed from the automated operation span task (one from the NCC group and one from the ST group) because they achieved less than 85 % accuracy on the math tasks (which was a task requirement). The Time × Group ANOVA revealed that neither the main effect of Time, $F(1,61) = 1.44$, $p = 0.235$, $\eta_p^2 = 0.02$, nor the main effect of Group, $F(2,61) = 2.72$, $p = 0.074$, $\eta_p^2 = 0.08$, nor the Interaction, $F(2,61) = 1.93$, $p = 0.155$, $\eta_p^2 = 0.06$, was significant. For the automated symmetry span task, the Time × Group ANOVA revealed a significant main effect of Time, $F(1,63) = 7.90$, $p = 0.007$, $\eta_p^2 = 0.11$; but neither the main effect of Group, $F(2,63) = 2.34$, $p = 0.105$, $\eta_p^2 = 0.07$, nor the Interaction, $F(2,63) = 2.33$, $p = 0.105$, $\eta_p^2 = 0.07$, was significant. Therefore, as in Experiment 1, *n*-back training improvement did not transfer to either of our WM measures.

General fluid intelligence tasks: training-related performance improvement on *Gf* was measured by comparing the total number of correct items from BS1 to the total number of correct items from BS2 for both the RAPM task and the CCF task. For RAPM, the ANOVA revealed that neither the main effect of Time, $F(1,63) = 0.08$, $p = 0.776$, $\eta_p^2 = 0.001$, the main effect of Group, $F(2,63) = 0.09$,

Table 2 Difference scores for battery session 1 and battery session 2

	Battery session 2–battery session 1; difference score		
	Control group	Spatial training group	Verbal training group
Working memory			
Automated operation span	−2.4 (15.8)	5.1 (11.8)	3.1 (10.7)
Automated symmetry span	1.0 (11.6)	2.3 (7.5)	6.6 (8.4)
General fluid intelligence			
Raven's advanced progressive matrices	−0.36 (2.1)	−0.23 (2.5)	0.91 (3.1)
Cattell's culture fair	−0.23 (4.0)	0.82 (2.8)	−0.18 (2.4)
Attentional control			
Flanker[a]	2.5 (25.3)	14.0 (21.9)	24.0 (35.0)
Antisaccade[a]	43.8 (325.6)	200.6 (167.5)	161.7 (163.2)

[a] Reverse scoring

$p = 0.917$, $\eta^2_p = 0.003$, nor the Interaction, $F(2,63) = 1.68$, $p = 0.195$, $\eta^2_p = 0.05$, was significant. Similarly, for the CCF neither the main effect of Time, $F(1,63) = 0.26$, $p = 0.615$, $\eta^2_p = 0.004$, the main effect of Group, $F(2,63) = 1.85$, $p = 0.165$, $\eta^2_p = 0.06$, nor the Interaction, $F(2,63) = 0.65$, $p = 0.528$, $\eta^2_p = 0.02$, was significant. As in Experiment 1, in this experiment, Gf did not improve following WM training.

Attentional control tasks: training-related performance improvement on AC processes was evaluated by comparing BS1 performance to BS2 performance on both the flanker task and the antisaccade task. For the flanker task, five participants were removed from the analysis (one from the VT group, two from the ST group, and two from the NCC group) because performance (either accuracy or RT) deviated greater than 2 standard deviations from the mean on BS1. RTs on congruent trials were subtracted from RTs on incongruent trials to obtain a difference score representing the amount of interference between the two conditions; this served as the dependent variable. The Time × Group ANOVA revealed that the main effect of Time, $F(1,58) = 13.93$, $p < 0.001$, $\eta^2_p = 0.19$. The main effect of Group, $F(2,58) = 0.15$, $p = 0.861$, $\eta^2_p = 0.005$, was not significant and the Interaction, $F(2,58) = 3.07$, $p = 0.054$, $\eta^2_p = 0.10$, approached significance. For the antisaccade task, two participants were removed from the analysis (one from the ST group and one from the NCC group) because their RTs were greater than two standard deviations slower than the mean on BS1. RTs for correct trials were assessed to measure performance. Again, the ANOVA revealed that the main effect of Time, $F(1,61) = 22.06$, $p < 0.001$, $\eta^2_p = 0.27$. Again the main effect of Group, $F(2,61) = 0.09$, $p = 0.918$, $\eta^2_p = 0.003$, was not significant and the Interaction, $F(2,61) = 2.64$, $p = 0.079$, $\eta^2_p = 0.08$, approached significance. In this experiment, there was a near-significant trend toward AC improvement following *n*-back training.

Discussion

Three main conclusions can be drawn from these data. First, adaptive *n*-back training can be effective in improving performance on some, but not all, untrained tasks (e.g., VSTM); furthermore, transfer appears to only occur when the processes that improved during training are also required for the transfer tasks (e.g., Dahlin et al., 2008a, b; Melby-Lervåg & Hulme, 2013). Second, WM training is an effective means of improving performance on the change detection task; and training influences both number and resolution processes. To the best of our knowledge, this is the first study demonstrating significant improvement on a measure of VSTM number and resolution following WM training. Interestingly, training did not have a similar beneficial effect on VSTM as measured via the short-term recall task. This suggests that compared to the short-term recall task, the change detection task may be a more sensitive measure of number and resolution processes and their interaction with cognitive training. Finally, ensuring variability on storage demands during training does, in fact, promote a high level of performance across training trials. This may be important if the amount of improvement demonstrated during training is related to the amount of improvement on other untrained tasks (see "General discussion").

Contact control group

A limitation of Experiment 1 was the lack of contact control group; therefore in Experiment 2, two experimental groups were designed each to serve as the contact control group for the other. While selecting highly similar training tasks is advantageous in that, the tasks are matched on both task requirements and difficulty, because the underlying processes are very similar, the likelihood of identifying different training effects between the groups may be reduced. In Experiment 2, a spatial and a verbal version of the adaptive *n*-back task were selected with the hypothesis that there might be stimulus-specific benefits of training on some of the battery tasks; however, these groups produced similar results. Despite this similarity, we do not believe that these data can be consistently interpreted as a consequence of a Hawthorne effect, demand characteristic, or group motivation argument (cf. Shipstead et al., 2010). In the present data, there are not universal benefits of training on all tasks measured. It seems unlikely that differences in motivation, expectations, or demand characteristics would have a differential effect on the various battery of tasks used here. Furthermore, it is perhaps also useful to note that a recent study with children directly assessing the difference between active and passive (no-contact) control groups, no differences were identified (Thorell, Lindqvist, Nutley & Klingberg, 2009) providing an early indication that NCC groups may provide a reliable comparison group. While we do not intend to undermine the importance that motivation or expectation may play a strong role in training studies, we point out that in the current data, there is no obvious reason to believe that motivation or expectation works selectively among the battery tasks.

General discussion

The purpose of the present experiments was to evaluate the efficacy of WM training on a variety of cognitive functions with a particular focus on VSTM (a previously under-investigated process in the cognitive training literature). The present data are consistent with two previous studies reporting a positive effect of training on VSTM (Kundu et al., 2013; Owens et al., 2013). The current study extends these findings by evaluating the effect of training on number and resolution subprocesses of VSTM. The data provide unique insights into cognitive training and VSTM specifically as well as add to the growing cognitive training literature to help paint a coherent picture of cognitive training efficacy more generally.

Transfer of cognitive skill

Visual short-term memory

Recent cognitive training investigations have reported a positive effect of training on VSTM capacity in both healthy adults (Kundu et al., 2013) as well as dysphoric patients (Owens et al., 2013). In both of these studies, the change detection task was used to evaluate VSTM. The current study sought to extend these findings by investigating the role of WM training of VSTM more rigorously both by including multiple measures of VSTM and by specifically investigating the effect of training on two subprocesses of VSTM, namely number and resolution.

Experiments 1 and 2 both showed a significant effect of training on the change detection task. Furthermore, in Experiment 2, individual capacity estimates were extracted from the change detection task data revealing a significant increase in capacity following adaptive *n*-back training for the ST and VT groups, but not the NCC group. These data replicate previous findings that VSTM capacity can improve with training (Kundu et al., 2013; Owens et al., 2013).

To the best of our knowledge, this is the first study that has directly assessed the influence of WM training on the individual subcomponents of VSTM (i.e., number and resolution). Experiment 1 showed a significant effect of WM training improvement on VSTM resolution and a trend for an effect on VSTM number. Experiment 2 modified the training design to promote variable practice and increased improvement during training. The Experiment 2 change detection data demonstrated that brain training significantly improved both measures of number and resolution for both training groups compared to the NCC group. No training-related improvements were evident for short-term recall task performance suggesting that while both the change detection task and visual short-term memory task are used to measure VSTM, perhaps the change detection task provides a more sensitive measure of training induced change.

Nevertheless, the change detection resolution finding is somewhat surprising. Past research suggests that VSTM resolution can be altered, but only when participants have expertise with the stimuli (Curby & Gauthier, 2007; Scolari et al., 2008). In both experiments, all groups had equal exposure to the specific stimuli; therefore, the training groups should not have an elevated levels of expertise compared to the NCC groups. However, improvement is evident for all training groups in both experiments. A proposed mechanism for improvement is outlined below.

Other cognitive processes

Much of the cognitive training literature has focused on the impact of cognitive training on cognitive processes such as WM, Gf, and AC with mixed results. For example, several studies report no effect of training on a variety of measures of WM (Jaeggi et al., 2008; Li et al., 2008; Redick et al., 2012; Schmiedek et al., 2010), while others report significant training-related improvements (Chein & Morrison, 2010). Similarly, many studies that assess the effectiveness of cognitive training on measures of Gf report significant post-training improvement (e.g., Colom et al. 2010; Jaeggi et al., 2008; Jaeggi et al., 2010b; Klingberg et al., 2002; Olesen et al., 2004; Rudebeck, Bor, Ormond, O'Reilly &

Lee, 2012; Jaušovec & Jaušovec, 2012), while many other studies fail to show an effect of training (e.g., Chein & Morrison, 2010; Dahlin et al., 2008b; Westerberg & Klingberg, 2007; Redick et al., 2012). Our data are most consistent with those experiments that suggest that *n*-back training does not improve untrained WM of Gf measures.

One of the most consistent findings in the cognitive training literature is that cognitive training transfers to measures of AC (Chein & Morrison, 2010; Klingberg et al., 2002; Olesen et al., 2004; Westerberg & Klingberg, 2007; see Owen et al., 2010; Dahlin et al., 2008a for exceptions); however, in all cases only the Stroop task was used. Our Experiment 1 is unique in that AC was assessed using multiple real-world military specific measures of AC. Training-related improvements in AC were not evident using these measures, and more traditional psychological measures of AC were used in Experiment 2. Experiment 2 is unique in that AC was assessed using both the flanker task and the antisaccade task. Training-related improvement on both the flanker and antisaccade tasks approached significance. These tasks were selected because of their high and shared loadings onto an AC construct (e.g., Unsworth & Spillers, 2010). To further investigate the impact of cognitive training on AC mechanistically, z-scores were calculated and composite measures were extracted for BS1 and BS2. And composite scores were submitted to a Time (BS1 vs. BS2) × Group (NCC vs. ST vs. VT) repeated measures ANOVA. Neither the main effect of Time, $F(1,56) = 0.05$, $p = 0.826$, $\eta_p^2 = 0.001$, nor the main effect of Group, $F(2,56) = 0.08$, $p = 0.926$, $\eta_p^2 = 0.003$, was significant. The Interaction, $F(2,56) = 3.86$, $p = 0.027$, $\eta_p^2 = 0.12$, was significant. Post hoc analysis revealed that the both the VT, $t(39) = 2.33$, $p = 0.025$, and ST, $t(36) = 2.31$, $p = 0.027$, groups performed better after training compared to the NCC group. There was no difference between the ST and VT groups, $t(37) = 0.24$, $p = 0.811$. While further research is necessary to investigate the effects of WM training on these AC tasks specifically, data add to this growing body of literature by demonstrating training-related improvement on a composite measure of AC.

The training task

Critical evaluation of the training data from Experiment 1 revealed that while an adaptive training design was used to promote task variability, this goal was not optimized due to stagnant levels of moderate performance. The importance of variability during training has been highlighted in both the skill learning and cognitive training literatures (c.f. Morrison & Chein, 2011; Schmidt & Bjork, 1992) and for this reason, the standard adaptive task requirements (e.g., Jaeggi et al., 2008; Jaeggi et al., 2010b; Redick et al., 2012) were modified in Experiment 2. Thus, Experiment 2 was unique in the literature in that it had the additional requirement that no individual could complete more than five blocks in a row at the same level of difficulty ensuring that once the participant reached a plateau, training difficulty would continue to vary. This design was effective in that every participant showed improvement (and in some cases, large improvements) across the eight training tasks. For the VT group, by training session 8, participants were able to perform at levels of *n* that were 3–12 (mean = 4.7 ± 2.2) higher than during session 1. Similarly, for the ST group, by training session 8, participants were able to perform at levels of *n* that were 1–8 (mean = 3.3 ± 1.6) higher than during session 1. Thus, it appears that the optimized design indeed facilitated training improvement across eight training sessions.

While neither experiment alone allows for a direct comparison between the optimized training design and a sub-optimal design, comparing the training data across experiments may be interesting to consider; therefore, we conducted Training Session (1–8) × Experiment (Experiment 1 vs. Experiment 2) repeated measures ANOVAs for the verbal task and the spatial task separately (Note: for this analysis, Experiment 2 data were reanalyzed to only include the first 18 blocks, as only 18 blocks of each task were performed in Experiment 1). The Interaction was significant for both the verbal, $F(6.2,291.5) = 5.58$, $p < 0.001$, $\eta_p^2 = 0.11$, and spatial, $F(5.2,244.9) = 2.36$, $p = 0.038$, $\eta_p^2 = 0.05$, tasks (Fig. 10b). Thus, for both the spatial and the verbal adaptive *n*-back tasks, participants in Experiment 2 demonstrated greater improvement across training than the participants in Experiment 1. While further research is necessary to confirm this finding, these data suggest that consistently varying training task difficulty could promote better learning and performance during training.

What is being trained?

The majority of studies in the cognitive training literature that train with the *n*-back task are categorized as WM training studies. However, while successful performance on the *n*-back task surely requires WM storage processes (e.g., Shipstead et al., 2010; McElree, 2001; Jaeggi et al., 2010a; Jonides et al., 1997), there are other cognitive processes engaged and the underlying neural mechanisms of WM training are not well understood (Buschkuehl, Jaeggi & Jonides, 2012). For example, AC processes are also essential for accurate performance on the *n*-back task (e.g., Jaeggi et al., 2008; Shipstead et al., 2010; Verhaeghen, Cerella & Basak, 2004; McElree, 2001). The current data suggest that both storage and AC are improved during training with the *n*-back task. We propose that it is

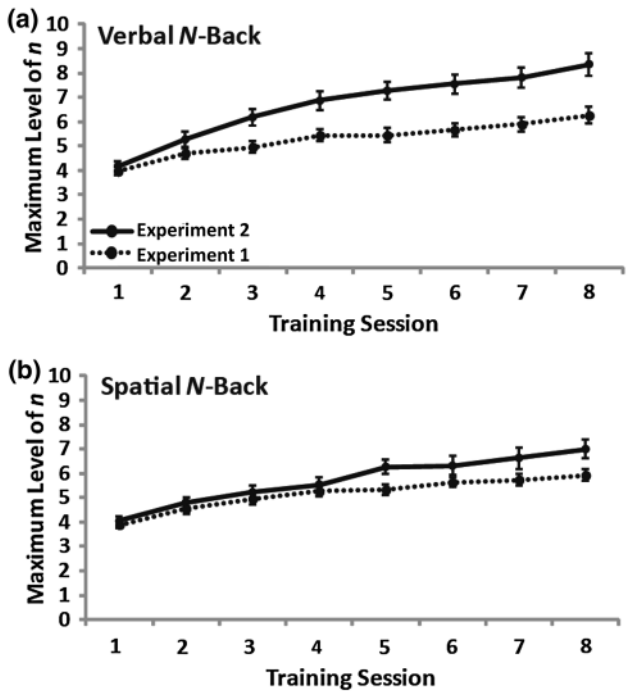

Fig. 10 Maximum level achieved on the **a** verbal and **b** spatial *n*-back tasks for Experiment 1 and Experiment 2

improved AC processes that are primarily responsible for improvement on the various battery tasks.

Attentional control is the process, or set of processes, that allows an individual to select task/situation relevant stimuli and ignore other stimuli (e.g., Neill, Valdes & Terry, 1994). In other words, AC facilitates relevant processing while inhibiting irrelevant processing to ensure optimal performance. Attention can select on certain physical attributes of the stimulus such as color and location (e.g., Neill et al., 1994; Broadbent, 1958). In the spatial version of the adaptive *n*-back task, individual stimuli are differentiated by their spatial location; however, equally critical is each stimulus' position in time. With practice during training, the ability to select on time and space is refined. In the verbal version of the adaptive *n*-back task, spatial location is not relevant; however, again temporal position is critical for successful performance. Equally important for performance success on both versions of the task is the ability to inhibit previous trial instructions in favor of current task goals. Thus, learning to inhibit non-task relevant information both within a trial and between trials is critical for successful performance on the adaptive *n*-back task. In fact, inhibitory control has been identified as the critical mechanism improved during WM training (Owens et al., 2013). Our data support this idea and are consistent with the proposition that training-related improvements to transfer tasks require mechanistic overlap (e.g., Dahlin et al., 2008a; Melby-Lervåg & Hulme, 2013; Barnett & Ceci, 2002). Therefore, in our study, we suggest that only those tasks that require inhibitory control improve following training.

Explanation of failed transfer

If this hypothesis holds merit, it follows that performance on the RAPM and CCF tasks, the motion interference and rapid decision-making tasks, as well as the short-term recall task would not show a benefit of *n*-back training given that in these tasks all of the stimuli are relevant to responding appropriately. In the RAPM task, to correctly identify the missing information, an individual must pay close attention to all of the other stimuli to try to understand what best completes the pattern. Inhibiting any part of the problem set is not useful because critical information about the best solution is embedded in all of the related stimuli. Furthermore, all of the relevant information for a given problem is available simultaneously, so temporal distinctions are not necessary. In the CCF task, similarly all of the information necessary to solve the problem is presented concurrently and all of the choices must be considered and compared to arrive at the appropriate solution. These tasks, therefore, rely little if at all on the component processes that were improved via single *n*-back training. Furthermore, in both of these tasks, the stimuli are unique on each trial so it is unlikely that considerable cross-trial interference occurs.

Similarly, in the rapid decision-making task, participants must consider all of the information (i.e., both the identity of the targets and their location in space) to respond accurately. Inhibiting any of the information is unnecessary and, in fact, detrimental to task performance. Additionally, the rule set remains constant across trials so understanding the relational hierarchy on one trial may aid performance on a subsequent trial. In the motion interference task, participants must respond to the letter task while keeping track of the trajectory of an invisible ball. Inhibiting the information on either task results in inaccurate response. Finally, in the short-term recall tasks, all stimuli in the memory set have an equally likely chance of being probed for retrieval, therefore, the best strategy is to select all available stimuli. Thus, improving the ability to inhibit irrelevant stimuli within a trial is not beneficial to task performance, and consequently performance would likely not benefit from adaptive *n*-back training.

Explanation of successful transfer

Each of the tasks used to measure WM and AC as well as the change detection measure of VSTM share some common features. First, stimuli were consistent across trials and therefore previous trial stimuli likely interfered with current trial to some degree creating between-trial interference. Participants must inhibit those stimuli that were present in the previous trial and select only those stimuli relevant for the current trial to perform accurately. If *n*-back training indeed improves between-trial AC, then we would predict successful transfer to these tasks. And perhaps this is why we saw a hint of successful transfer to some WM measures (i.e., automated operation span in Experiment 1). Each of the complex span tasks used to measure WM required participants to keep track of information, which was temporally interleaved with irrelevant information. Thus, participants were required to keep track of some information (i.e., the letter in the automated operation span task and the spatial locations in the automated symmetry span task) while inhibiting irrelevant information (i.e., math problems in the automated operation span task and symmetry judgments in the automated symmetry span task) presented just before and after that relevant information. Thus, participants were required to ignore irrelevant information separated from relevant information in time.

However, we have proposed that *n*-back training improves both between-trial and within-trial attentional control, and indeed flanker, antisaccade, and change detection tasks each require participants to inhibit irrelevant information within trials. In each of these tasks, on any given trial, there is relevant and irrelevant information presented simultaneously on the display. In the flanker tasks, participants must respond to the direction of a central arrow while ignoring the direction of the flanking arrows; thus the target is distinguishable from the distractors based solely on spatial location. In the antisaccade task, on each trial, participants must ignore information on one side of the display and attend to information on the other side of the display distinguishable only by a cue. Finally, in the change detection task, participants are cued to one side of the display and are responsible for remembering the stimuli that appear on that side of space. Inhibiting the stimuli on the uncued side of space is essential to accurate performance. Thus, if it is an individual's ability to attend to relevant information and inhibit irrelevant information that is honed during *n*-back training, then we would indeed predict post-training improvement on each of these tasks.

Conclusions

General conclusions from this set of experiments are twofold. First, VSTM capacity can be improved following cognitive training and that improvement is evident both in the number of items that can be held in memory, and the resolution with which those items are distinguishable. Second, we propose that VSTM improves because it relies on inhibitory mechanisms that are likely enhanced during *n*-back training. Consequently, other cognitive processes that require inhibitory control (e.g., AC) also benefit from *n*-back training; those processes that do not rely on inhibition (e.g., Gf) do not show a benefit.

Acknowledgments This research was funded in part by a grant to the author from the Air Force Offices of Scientific Research *(Award No. FA9550-09-1-0162)*; in part by a grant from the American Psychological Association; and in part by a graduate research award from the Georgia Institute of Technology School of Psychology.

References

Adams, J. A. (1987). Historical review and appraisal of research on the learning, retention, and transfer of human motor skills. *Psychological Bulletin, 101*(1), 41–74. doi:10.1037/0033-2909.101.1.41.

Arend, A. M., & Zimmer, H. D. (2012). Successful training of filtering mechanisms in multiple object tracking does not transfer to filtering mechanisms in a visual working memory task: behavioral and electrophysiological evidence. *Neuropsychologia, 50*(10), 2379–2388. doi:10.1016/j.neuropsychologia.2012.16.007.

Awh, E., Barton, B., & Vogel, E. K. (2007). Visual working memory represents a fixed number of items regardless of complexity. *Psychological Science, 18*(7), 622–628. doi:10.1111/j.1467-9280.2007.01949.x.

Bahrick, H. P. (1984). Semantic memory content in permastore: fifty years of memory for spanish learned in school. *Journal of Experimental Psychology: General, 113*(1), 1–29.

Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn? A taxonomy for far transfer. *Psychological Bulletin, 128*(4), 612–637. doi:10.1037//0033-2909.128.4.612.

Bartlett, F. C. (1947). The measurement of human skill. *British Medical Journal, 1*(4511), 877–880.

Boot, W. R., Kramer, A. F., Simons, D. J., Fabiani, M., & Gratton, G. (2008). The effects of video game playing on attention, memory, and executive control. *Acta Psychologica, 129*(3), 387–398. doi:10.1016/j.actpsy.2008.09.005.

Brehmer, Y., Westerberg, H., & Bäckman, L. (2012). Working-memory training in younger and older adults: training gains, transfer, and maintenance. *Frontiers in Human Neuroscience,*. doi:10.3389/fnhum.2012.00063.

Broadbent, D. E. (1958). *Perception and communication*. New York: Pergamon Press.

Buschkuehl, M., Jaeggi, S. M., & Jonides, J. (2012). Neuronal effects following working memory training. *Developmental Psychology, 2*, S167–S179. doi:10.1016/j.dcn.2011.10.001.

Cattell, R. B. (1949). *Culture free intelligence test, scale 1, Handbook*. Champaign: Institute of Personality and Ability.

Chein, J. M., & Morrison, A. B. (2010). Expanding the mind's workspace: training and transfer effects within a complex working memory span task. *Psychonomic Bulletin & Review, 17*(2), 193–199. doi:10.3758/PBR.17.2.193.

Colom, R., Quiroga, M. A., Shih, P. C., Martinez, K., Burgaleta, M., Martinez-Molina, A., et al. (2010). Improvement in working memory is not related to increased intelligence scores. *Intelligence, 38*, 497–505. doi:10.1016/j.intell.2010.08.002.

Conway, A. R. A., Cowan, N., & Bunting, M. F. (2001). The cocktail party phenomenon revisited: the importance of working memory

capacity. *Psychonomic Bulletin & Review, 8*(2), 331–335. doi:10.3758/BF03196169.
Cowan, N. (2001). The magical number 4 in short-term memory: a reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24*, 87–185.
Curby, K. M., & Gauthier, I. (2007). A visual short-term memory advantage for faces. *Psychonomic Bulletin & Review, 12*, 1127–1133. doi:10.3758/BF03196811.
Dahlin, E., Neely, A. S., Larsson, A., Backman, L., & Nyberg, L. (2008a). Transfer of learning after updating training mediated by the striatum. *Science, 320*, 1510–1512. doi:10.1126/science.1155466.
Dahlin, E., Nyberg, L., Backman, L., & Neely, A. S. (2008b). Plasticity of executive functioning in young and older adults: immediate training gains, transfer, and long-term maintenance. *Psychology and Aging, 23*(4), 720–730. doi:10.1037/a0014296.
Dempster, F. N. (1988). The spacing effect: a case study in the failure to apply the results of psychological research. *American Psychologist, 43*(8), 627–634. doi:10.1037//0003-066X.43.8.627.
Ebbinghaus, H. E. (1885/1913). *Memory: A contribution to experimental psychology*. New York: Dover.
Engle, R. W. (2001). What is working-memory capacity? In H. L. Roediger & J. S. Nairne (Eds.), *Nature of Remembering: Essays in Honor of Robert G. Crowder* (pp. 297–314). Washington: American Psychological Association.
Engle, R. W., Kane, M. J., & Tuholski, S. W. (1999). Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence, and functions of the prefrontal cortex. In A. Miyake & P. Shah (Eds.), *Models of working memory: mechanisms of active maintenance and executive control* (pp. 102–134). Cambridge: Cambridge University Press.
Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception and Psychophysics, 16*(1), 143–149. doi:10.3758/BF03203267.
Fukuda, K., & Vogel, E. K. (2009). Human variation in overriding attentional capture. *Journal of Neuroscience, 29*(27), 8726–8733. doi:10.1523/JNEUROSCI.2145-09.2009.
Fukuda, K., & Vogel, E. K. (2011). Individual differences in recovery time from attentional capture. *Psychological Science, 22*(3), 361–368. doi:10.1177/0956797611398493.
Fukuda, K., Vogel, E. K., Mayr, U., & Awh, E. (2010). Quantity not quality: the relationship between fluid intelligence and working memory capacity. *Psychonomic Bulletin & Review, 17*(5), 673–679. doi:10.3758/17.5.673.
Gaspar, J. G., Neider, M. B., Simons, D. S., McCarley, J. S., & Kramer, A. F. (2013). Change detection: training and transfer. *PLoS One, 8*(6), e67781. doi:10.1371/journal.pone.0067781.
Green, C. S., & Bavelier, D. (2003). Action video game modifies visual selective attention. *Nature, 423*, 534–537. doi:10.1038/nature01647.
Green, C. S., & Bavelier, D. (2006). Effect of action video games on the spatial distribution of visual attention. *Journal of Experimental Psychology: Human Learning and Memory, 32*(6), 1465–1478. doi:10.1037/0096-1523.32.6.1465.
Heitz, R. P., & Engle, R. W. (2007). Focusing the spotlight: individual differences in visual attention control. *Journal of Experimental Psychology: General, 136*(2), 217–240. doi:10.1037/0096-3445.136.2.217.
Jaeggi, S. M., Buschkuehl, M., Jonides, J., & Perrig, W. J. (2008). Improving fluid intelligence with training on working memory. *Proceedings of the National Academy of Sciences, 105*(19), 6829–6833. doi:10.1073/pnas.0801268105.
Jaeggi, S. M., Buschkuehl, M., Perrig, W. J., & Meier, B. (2010a). The concurrent validity of the *n*-back task as a working memory meausre. *Memory, 18*(4), 394–412.
Jaeggi, S. M., Studer-Luethi, B., Buschkuehl, M., Su, Y.-F., Jonides, J., & Perrig, W. J. (2010b). The relationship between *n*-back performance and matrix reasoning—implications for training and transfer. *Intelligence, 38*(6), 625–635. doi:10.1016/j.intell.2010.09.001.
Jaeggi, S. M., Buschkuehl, M., Jonides, J., & Shah, P. (2011). Short- and long-term benefits of cognitive training. *Proceedings of the National Academy of Sciences, 108*(25), 11081–11086. doi:10.1073/pnas.1103228108.
Jaušovec, N., & Jaušovec, K. (2012). Working memory training: improving intelligence—Changing brain activity. *Brain and Cognition, 79*, 96–106. doi:10.1016/j.bandc.2012.02.007.
Jonides, J., Schumacher, E. H., Smith, E. E., Lauber, E. J., Awh, E., Minoshima, S., & Koeppe, R. A. (1997). Verbal working memory load affects regional brain activation as measured by PET. *Journal of Cognitive Neuroscience, 9*(4), 462–475.
Kane, M. J., Brown, L. H., McVay, J. C., Silvia, P. J., Myin-Germeys, I., & Kwapil, T. R. (2007a). For whom the mind wanders, and when: an experience-sampling study of working memory and executive control in daily life. *Psychological Science, 18*, 614–621. doi:10.1111/j.1467-9280.2007.01948.x.
Kane, M. J., Conway, A. R. A., Miura, T. K., & Colflesh, G. J. H. (2007b). Working memory, attention control, and the *n*-back task: a question of construct validity. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 33*(3), 615–622. doi:10.1037/0278-7393.33.3.615.
Kane, M. J., & Engle, R. W. (2003). Working-memory capacity and the control of attention: the contributions of goal neglect, response competition, and task set to Stroop interference. *Journal of Experimental Psychology: General, 132*(1), 47–70. doi:10.1037/0096-3445.132.1.47.
Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The domain generality of working-memory capacity: a latent-variable approach to verbal and spatial memory span and reasoning. *Journal of Experimental Psychology: General, 133*, 189–217. doi:10.1037/0096-3445.133.2.189.
Kerr, R., & Booth, B. (1978). Specific and varied practice of motor skill. *Perceptual and Motor Skills, 46*(2), 395–401.
Klingberg, T., Forssberg, H., & Westerberg, H. (2002). Increased brain activity in frontal and parietal cortex underlies the development of visuospatial working memory capacity during childhood. *Journal of Cognitive Neuroscience, 14*(1), 1–10. doi:10.1162/089892902317205276.
Kundu, B., Sutterer, D. W., Emrich, S. M., & Postle, B. R. (2013). Strengthened effective connectivity underlies transfer of working memory training to tests of short-term memory and attention. *The Journal of Neuroscience, 33*(20), 8705–8715. doi:10.1523/JNEUROSCI.5565-12.2013.
Li, S. C., Schmiedek, F., Huxhold, O., Rocke, C., Smith, J., & Lindenberger, U. (2008). Working memory plasticity in old age: practice gain, transfer, and maintenance. *Psychology and Aging, 23*(4), 731–742. doi:10.1037/a0014343.
Lilienthal, L., Tamez, E., Shelton, J. T., Myerson, J., & Hale, S. (2013). Dual *n*-back training increases the capacity of the focus of attention. *Psychonomic Bulletin & Review, 20*, 135–141. doi:10.3758/s12423-012-0335-6.
McDonald, J. H. (2008). *Biological Statistics*. Baltimore: Sparky House Publishing.
McElree, B. (2001). Working memory and focal attention. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 27*, 817–835. doi:10.1037//0278-7393.27.3.817.
Melby-Lervåg, M., & Hulme, C. (2013). Is working memory training effective? *A meta-analytic review. Developmental Psychology, 49*(2), 270–291. doi:10.1037/a0028228.
Morrison, A. B., & Chein, J. M. (2011). Does working memory training work? The promise and challenges of enhancing

cognition by training working memory. *Psychological Bulletin & Review, 18*, 46–60. doi:10.3758/s13423-010-0034-0.
Morrison, A. B., & Chein, J. M. (2012). The controversy over Cogmed. *Journal of Applied Research in Memory and Cognition, 1*(3), 208–210. doi:10.1016/j.jarmac.2012.07.005.
Motulsky, H. (2010). Multiple comparisons concepts. In *Intuitive Biostatistics: A Nonmathematical Guide to Statistical Thinking* (Vol. 2, pp. 157–168). New York: Oxford University Press, Inc.
Neill, W., Valdes, L., & Terry, K. (1994). Selective attention and the inhibitory control of cognition. In F. N. Dempster & C. Brainerd (Eds.), *New perspectives on interference and inhibition in cognition*. New York: Academic Press.
O'Donnell, R. D., Moise, S., & Schmidt, R. M. (2005). Generating performance test batteries relevant to specific operational tasks. *Aviation, Space and Environmental Medicine, 76*(7), C24–C30.
Olesen, P. J., Westerberg, H., & Klingberg, T. (2004). Increased prefrontal and parietal activity after training of working memory. *Nature Neuroscience, 7*(1), 75–79. doi:10.1038/nn1165.
Owen, A. M., Hampshire, A., Grahn, J. A., Stenton, R., Dajani, S., Burns, A. S., et al. (2010). Putting brain training to the test. *Nature, 465*(7299), 775–779. doi:10.1038/nature09042.
Owens, M., Koster, E. H. W., & Derakshan, N. (2013). Improving attentional control in dysphoria through cognitive training: transfer effects on working memory capacity and filtering efficiency. *Psychophysiology, 50*(3), 297–307. doi:10.1111/psyp.12010.
Pashler, H., Rohrer, D., Cepeda, N. J., & Carpenter, S. K. (2007). Enhancing learning and retarding forgetting: choices and consequences. *Psychonomic Bulletin & Review, 14*(2), 187–193. doi:10.3758/BF03194050.
Redick, T. S., Shipstead, Z., Harrison, T. L., Hicks, K. L., Fried, D. E., Hambrick, D. Z., et al. (2012). No evidence of intelligence improvement after working memory training: a randomized, placebo-controlled study. *Journal of Experimental Psychology: General, 142*(2), 359–379. doi:10.1037/a0029082.
Rogers, W. A. (1996). Assessing age-related differences in the long-term retention of skills. In W. A. Rogers, A. D. Fisk, & N. Walker (Eds.), *Aging and skilled performance: Advances in theory and application* (pp. 185–200). Hillsdale: Erlbaum.
Rudebeck, S. R., Bor, D., Ormond, A., O'Reilly, J. X., & Lee, A. C. H. (2012). A potential spatial working memory training task to improve both episodic memory and fluid intelligence. *PLoS One, 7*(11), e50431. doi:10.1371/journal.pone.0050431.
Schmidt, R. A., & Bjork, R. A. (1992). New conceptualizations of practice: paradigms suggest new concepts for training. *Psychological Science, 3*(4), 207–217. doi:10.1111/j.1467-9280.1992.tb00029.x.
Schmiedek, F., Lövdén, M., & Lindenberger, U. (2010). Hundred days of cognitive training enhance broad cognitive abilities in adulthood: findings from the COGITO study *Frontiers in Aging. Neuroscience, 2*(27), 1–10. doi:10.3389/fnagi.2010.00027.
Schneider, W., & Chen, J. M. (2003). Controlled and automatic processing: behavior, theory, and biological mechanisms. *Cognitive Science, 27*, 525–559.
Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime User's Guide*. Pittsburgh: Psychology Software Tools Inc.
Schneiders, J. A., Opitz, B., Krick, C. M., & Mecklinger, A. (2011). Separating intra-modal and across-modal training effects in visual working memory: an fMRI investigation. *Cerebral Cortex, 21*(11), 2555–2564. doi:10.1093/cercor/bhr037.
Schneiders, J. A., Opitz, B., Tang, H., Deng, Y., Xie, C. X., Li, H., et al. (2012). The impact of auditory working memory training on the fronto-parietal working memory network. *Frontiers in Human Neuroscience, 6*, 173. doi:10.3389/fnhum.2012.00173.
Scolari, M., Vogel, E. K., & Awh, E. (2008). Perceptual expertise enhances the resolution but not the number of representations in working memory. *Psychonomic Bulletin & Review, 15*(1), 215–222. doi:10.3758/pbr.15.1.215.
Shea, J. B., & Morgan, R. L. (1979). Contextual interference effects on the acquisition, retention, and transfer of a motor skill. *Journal of Experimental Psychology: Human Learning and Memory, 5*(2), 179–187. doi:10.1037//0278-7393.5.2.179.
Shipstead, Z., Redick, T. S., & Engle, R. W. (2010). Does working memory training generalize? *Psychologica Belgica, 50*, 245–276.
Thorell, L. B., Lindqvist, S., Nutley, S. B., & Klingberg, T. (2009). Training and transfer effects of executive functions in preschool children. *Developmental Science, 12*(1), 106–113. doi:10.1111/j.1467-7687.2008.00745.x.
Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language, 28*, 127–154. doi:10.1016/0749-596X(89)90040-5.
Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. [Research Support, U.S. Gov't, Non-P.H.S.]. *Behavior Research Methods, 37*(3), 498–505. doi:10.3758/BF03192720.
Unsworth, N., Schrock, J. C., & Engle, R. W. (2004). Working memory capacity and the antisaccade task: individual differences in voluntary saccade control. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 30*(6), 1302–1321. doi:10.1037/0278-7393.30.6.1302.
Unsworth, N., & Spillers, G. J. (2010). Working memory capacity: attention control, secondary memory, or both? A direct test of the dual-component model. *Journal of Memory and Language, 62*, 392–406. doi:10.1016/j.jml.2010.02.001.
Verhaeghen, P., Cerella, J., & Basak, C. (2004). A working memory workout: how to expand the focus of serial attention from one to four items in 10 hours or less. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 10*(6), 1322–1337. doi:10.1037/0278-7393.30.6.1322.
Vogel, E. K., McCollough, A. W., & Machizawa, M. G. (2005). Neural measures reveal individual differences in controlling access to working memory. *Nature, 438*(7067), 500–503. doi:10.1038/nature04171.
Westerberg, H., & Klingberg, T. (2007). Changes in cortical activity after training of working memory—a single-subject analysis. *Physiology & Behavior, 92*, 186–192. doi:10.1016/j.physbeh.2007.05.041.
Willis, S. L., Tennstedt, S. L., Marsiske, M., Ball, K., Elias, J., Koepke, K. M., et al. (2006). Long-term effects of cognitive training on everyday functional outcomes in older adults. *Journal of the American Medical Association, 296*(23), 2805–2814. doi:10.1001/jama.296.23.2805.
Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature, 453*(7192), 233–235. doi:10.1038/nature06860.
Zhang, W., & Luck, S. J. (2009). Sudden death and gradual decay in visual working memory. *Psychological Science*,. doi:10.1111/j.1467-9280.2009.02322.x.
Zhang, W., & Luck, S. J. (2011). The number and quality of representations in working memory. *Psychological Science, 22*(11), 1434–1441. doi:10.1177/0956797611417006.
Zimmer, H. D., Popp, C., Reith, W., & Krick, C. M. (2012). Gains of item-specific training in visual working memory and their neural correlates. *Brain Research, 1466*, 44–55. doi:10.1016/j.brainres.2012.05.019.