N-back versus Complex Span Working Memory Training

Kara J. Blacker[1], **Serban Negoita**[1], **Joshua B. Ewen**[1,2,3], and **Susan M. Courtney**[1,4,5]

[1]Department of Psychological & Brain Sciences, Johns Hopkins University

[2]Neurology and Developmental Medicine, Kennedy Krieger Institute

[3]Department of Neurology, Johns Hopkins University School of Medicine

[4]Department of Neuroscience, Johns Hopkins University School of Medicine

[5]F.M. Kirby Center for Functional Neuroimaging, Kennedy Krieger Institute

Abstract

Working memory (WM) is the ability to maintain and manipulate task-relevant information in the absence of sensory input. While its improvement through training is of great interest, the degree to which WM training transfers to untrained WM tasks (near transfer) and other untrained cognitive skills (far transfer) remains debated and the mechanism(s) underlying transfer are unclear. Here we hypothesized that a critical feature of dual n-back training is its reliance on maintaining relational information in WM. In Experiment 1, using an individual differences approach, we found evidence that performance on an n-back task was predicted by performance on a measure of relational WM (i.e., WM for vertical spatial relationships independent of absolute spatial locations); whereas the same was not true for a complex span WM task. In Experiment 2, we tested the idea that reliance on relational WM is critical to produce transfer from n-back but not complex span task training. Participants completed adaptive training on either a dual n-back task, a symmetry span task, or on a non-WM active control task. We found evidence of near transfer for the dual n-back group; however, far transfer to a measure of fluid intelligence did not emerge. Recording EEG during a separate WM transfer task, we examined group-specific, training-related changes in alpha power, which are proposed to be sensitive to WM demands and top-down modulation of WM. Results indicated that the dual n-back group showed significantly greater frontal alpha power after training compared to before training, more so than both other groups. However, we found no evidence of improvement on measures of relational WM for the dual n-back group, suggesting that near transfer may not be dependent on relational WM. These results suggest that dual n-back and complex span task training may differ in their effectiveness to elicit near transfer as well as in the underlying neural changes they facilitate.

Keywords

cognitive training; working memory; transfer; alpha power

Corresponding Author: Kara J. Blacker, Ph.D. Psychological and Brain Sciences, Johns Hopkins University, 143 Ames Hall, 3400 N Charles St. Baltimore, MD 21218, Phone: 410-516-5649, Fax: 410-516-4478, kblacker4@gmail.com.

In the past 15 years, a body of research on the effects of cognitive training has amassed. The overarching goal of cognitive training is to improve specific or broad-ranging cognitive skills and numerous approaches have been examined with this goal in mind, such as video game training (Green, Gorman, & Bavelier, 2016; Powers, Brooks, Aldrich, Palladino, & Alfieri, 2013), working memory (WM) training (Au et al., 2015; Morrison & Chein, 2011; Shipstead, Redick, & Engle, 2012), meditation training (Chiesa, Calati, & Serretti, 2011; Lutz, Slagter, Dunne, & Davidson, 2008; Tang, Holzel, & Posner, 2015), non-invasive brain stimulation (Berryhill, Peterson, Jones, & Stephens, 2014; Parkin, Ekhtiari, & Walsh, 2015), aerobic exercise (Hillman, Erickson, & Kramer, 2008) and many others (for a comprehensive review see, Simons et al., 2016). Of these approaches, WM training has, arguably, received the most attention and investigation, but also been surrounded by the most controversy.

The seminal study by Jaeggi, Buschkuehl, Jonides, and Perrig (2008) demonstrated that training on a WM task yielded improvement on a measure of fluid intelligence (*Gf*). While this result has been replicated many times (e.g., Jaeggi, Buschkuehl, Shah, & Jonides, 2014; Jausovec & Jausovec, 2012; Rudebeck, Bor, Ormond, O'Reilly, & Lee, 2012; Stephenson & Halpern, 2013; for a meta-analysis see Au et al. 2015), others have failed to replicate this effect (e.g., Redick et al., 2013; Thompson et al., 2013; for a review see Shipstead, Redick, & Engle 2012). Specifically, this area of investigation has focused on "far" transfer, whereby transfer represents the degree of similarity between the trained task (e.g., dual n-back) and untrained tasks (e.g., matrix reasoning tasks). Some of the debate in the WM training literature has surrounded the methods used to draw conclusions about transfer. For example, many early WM training studies, like Jaeggi et al. (2008) utilized passive control groups that did not engage in any training and were simply tested on the transfer measures twice. It is now widely accepted that making any firm conclusion about transfer effects requires an active control group, which controls for expectations, contact with experimenters, and other factors (see Simons et al., 2016 for a review of best practices). Another early criticism of the cognitive training field was the issue of inadequate baselines, whereby transfer might be concluded in the face of the control group performing worse after training (for a detailed consideration of this issue see, Boot et al., 2013). Furthermore, a critical feature of effective training that has evolved with the field is the use of adaptive training programs. For example, Lovden and colleagues (2010) argue that a fundamental prerequisite for successful cognitive training is a mismatch or imbalance between environmental demands and actual brain supply. To create this prolonged mismatch, training tasks must be challenging but manageable with a high degree of effort. This balance between keeping a task difficult enough to ensure the participant is not bored while easy enough to ensure the participant does not get frustrated is typically achieved through adaptive paradigms that keep the effort and feasibility level consistent for each individual participant.

Further, a source of heterogeneity in this literature that may account for some of the inconsistent results is the diverse set of WM tasks that have been used for training purposes. The most commonly used training task is a variant of the single or dual n-back task, but others have used complex span WM tasks (Kane et al., 2004; Redick et al., 2012), or commercialized programs like Cogmed (e.g., Holmes, Gathercole, & Dunning, 2009; Klingberg et al., 2005). Indeed, in our review of the literature we assessed that far transfer to

Gf was present most often in response to dual n-back training (Colom et al., 2013; Jaeggi et al., 2008; Jaeggi, Buschkuehl, Perrig, & Meier, 2010; Jaeggi et al., 2014; Jausovec & Jausovec, 2012), but there is no evidence of far transfer to *Gf* from complex span tasks (Chein & Morrison, 2010; Richmond, Morrison, Chein, & Olson, 2011). While there are examples of dual n-back training not resulting in far transfer to *Gf*, this type of training has shown the most promise in effectively improving *Gf* with training. Therefore, we sought to identify the "active ingredient" in the dual n-back that may make it more efficacious. Specifically, we aimed to directly compare two of the most commonly used tasks for both assessment and training of WM, namely an n-back task and a complex span task. N-back and complex span tasks, while both used to measure WM, have been shown to be only weakly correlated with one another (for a meta-analysis see, Redick & Lindsey, 2013). To the best of our knowledge, no previous study has directly compared n-back and complex span WM training in the same study.

Here we hypothesized that a critical difference between n-back and complex span tasks is their reliance on maintaining and updating relational information in WM (Jonides et al., 1997; Miyake et al., 2000). Matrix reasoning tasks that are typically used to assess *Gf* also involve extracting relational information between stimuli (Carpenter, Just, & Shell, 1990). Previous work from our group has demonstrated a neural dissociation between maintaining concrete and relational information in WM (Ackerman & Courtney, 2012; Blacker & Courtney, 2016; Blacker, Ikkai, Lakshmanan, Ewen, & Courtney, 2016; Ikkai, Blacker, Lakshmanan, Ewen, & Courtney, 2014). Specifically, these studies have shown that maintaining a concrete piece of sensory information, such as a spatial location, is supported by distinct neural substrates as compared to maintaining a spatial relationship that is independent of the original sensory location. Other groups have also shown dissociable neural substrates for processing or maintaining relational information from other types of concrete stimuli (Badre, 2008; Bahlmann, Blumenfeld, & D'Esposito, 2014; Libby, Hannula, & Ranganath, 2014). Based on this previous work, we hypothesized that the n-back task relies more on relational information than a complex span task does due to the constant need for spatiotemporal updating of information. It can be reasoned that n-back tasks may involve multiple forms of relational information depending on the task stimuli. For example, in a spatial n-back task, participants are asked to remember both the spatial relationship between a current and previously presented item, but also must remember the "temporal" relationship with respect to how many items have passed between any given two stimuli. While most n-back tasks likely involve some form of relational information, here we focused on visuospatial versions that seem to have multiple relations to maintain.

We first tested this hypothesis in a cross-sectional study examining individual differences in WM for spatial relations versus locations and performance on n-back and complex span tasks, as well as a measure of *Gf*. Next, we sought to test whether this differential reliance on relational information could explain the differences in training efficacy found between n-back and complex span tasks in the literature by examining near and far transfer effects in groups trained on either one task or the other. Further, the current study aimed to investigate the neural underpinnings of effective WM training by comparing the effects of these two training tasks on neural activity during performance of a spatial relation and location WM

task using EEG. In particular, we hypothesized that n-back training would have a greater effect on neural activity during relational WM than would complex span training.

In particular, we examined alpha power (8-13Hz) because a body of research has emerged suggesting oscillations in this range are involved in the maintenance of information in WM and have been linked to individual differences in WM performance (for a review see, Roux & Uhlhaas, 2013). Alpha band oscillations have been observed in a number of sensory modalities during WM maintenance (Haegens, Osipova, Oostenveld, & Jensen, 2010; Herrmann, Senkowski, & Rottger, 2004; Kaiser, Heidegger, Wibral, Altmann, & Lutzenberger, 2007). For example, it has recently been demonstrated that the topography of alpha tracks the spatial location of to-be-remembered items in WM (Foster, Sutterer, Serences, Vogel, & Awh, 2016). The majority of work has focused on alpha over posterior brain regions, which are thought to reflect the inhibition of task-irrelevant brain regions in an effort to prioritize processing of task-relevant information (Jensen, Gelfand, Kounios, & Lisman, 2002; Jokisch & Jensen, 2007; Kelly, Lalor, Reilly, & Foxe, 2006; Klimesch, Sauseng, & Hanslmayr, 2007). Moreover, WM performance has also been linked to alpha over frontal brain regions (Sadaghiani et al., 2012; Sauseng, Klimesch, Schabus, & Doppelmayr, 2005; Sauseng, Klimesch, Stadler, et al., 2005). Specifically, there is evidence that synchronization of alpha oscillations in posterior and frontal brain regions serves to protect items currently held in WM (Bonnefond & Jensen, 2012). In fact, Zanto, Rubens, Thangavel, and Gazzaley (2011) found that modulating this anterior-posterior alpha synchronization via rapid transcranial magnetic stimulation (rTMS) impaired WM performance, which suggests a causal role for prefrontal cortex (PFC) in modulating posterior brain regions. Taken together, this previous evidence suggests that alpha may represent a putative mechanism by which WM training may exert its effect on enhanced performance.

More specifically, in two related studies, our group has found that when participants maintain a spatial relation in WM, compared to a spatial location, there is an increase in posterior alpha power (Blacker et al., 2016; Ikkai et al., 2014). This increase in posterior alpha power has been interpreted as representing suppression of sensory brain regions because the sensory information (i.e., the spatial locations) is no longer task-relevant. In addition to differences in posterior alpha, our previous work has also shown that there is greater frontal alpha power when a relationship is being maintained in WM compared to a location, as well as increased frontal-posterior phase synchrony (Blacker et al., 2016). Therefore, if dual n-back and complex span differentially rely on relational WM, and if dual n-back training strengthens relational WM, then changes in alpha power may represent a neural marker of improvement that would be greater for dual n-back training. Thus in the current study, we investigated changes in alpha power during this same type of spatial relations and locations WM tasks before and after dual n-back and complex span WM training.

Here we describe the results of two experiments. Experiment 1 is a behavioral study testing our hypothesis that the n-back task relies more on relational WM as compared to complex span tasks. The results of that study supported the idea that the reliance on relational processing in the n-back task could account for differences in transfer following training that

have been reported in the literature. In Experiment 2, we aimed to directly compare the efficacy of dual n-back and complex span training to elicit near and far transfer and we aimed to investigate a potential neural basis of these training-induced changes in WM performance: changes in alpha power as measured by EEG.

Experiment 1: Individual Differences Study

Method

Participants—Twenty-four healthy adults (age: M=19.50, SD=1.2; 8 males) participated for course credit. All participants had normal or corrected-to-normal vision and reported no current and past neurological or psychological conditions (e.g., traumatic brain injury, major depression, epilepsy…etc), and gave written informed consent approved by the Institutional Review Board of Johns Hopkins University.

Tasks and Procedures—Experimental stimuli were controlled by MATLAB (The MathWorks, Natick, MA) using Psychophysics Toolbox extensions (Brainard, 1997; Pelli, 1997), and displayed on a 21" Apple iMac monitor. Participants were seated approximately 60cm from the monitor.

Participants completed one session, which included performance of the following tasks in one of two orders, counterbalanced across participants: Order 1) Spatial Locations and Relations WM task, Symmetry Span, single spatial n-back task, and BOMAT; Order 2) BOMAT, Spatial Locations and Relations WM task, single spatial n-back task, Symmetry Span.

Symmetry Span: Participants completed the automated Symmetry Span task (Kane et al., 2004). Participants recalled sequences of 2-5 red-square locations while performing an interleaved symmetry judgment task. The dependent measure for Symmetry Span was the partial score[1], which is the sum of red squares recalled in the correct location and serial order, regardless of whether the entire trial was recalled correctly. This score will henceforth be referred to as the Symmetry Span Score.

Spatial n-back: Participants were shown a sequence of spatial locations indicated by black squares and were asked to respond each time the current stimulus was in the same location as the one presented n positions back in the sequence. The black squares were presented in 8 possible locations organized around an imaginary circle around fixation. The squares were presented on a gray background for 500ms each, followed by a 2500ms interstimulus interval (ISI). Participants were required to press the SPACE bar for targets, and their response window lasted from the onset of the stimulus until the presentation of the next stimulus (3000ms); no response was required for non-targets. Participants were tested on 1-, 2-, 3-, 4-, and 5-back levels in that order. A level consisted of 20+n stimuli and contained 6 targets and 14+n non-targets each. For each n-level, we calculated d' as a measure of sensitivity (d'=Z(hit rate) – Z(false alarm rate)), which is a useful metric for an n-back task

---

[1]The Symmetry Span task traditionally yields two scores: "partial" and "absolute". Here we focus on partial scores as these have been shown to have higher internal consistency than absolute scores (e.g., Conway et al., 2005).

because it measures hit rate while penalizing for false alarms (e.g., Haatveit et al., 2010). We used the average d' across all 5 levels as the dependent measure of performance.

Bochumer Matrices Test (BOMAT): We assessed fluid intelligence (*Gf*) using the BOMAT (Hossiep, Turck, & Hasella, 1999). The BOMAT is a matrix-reasoning test that contains 29 problems and has two full versions, A and B. All participants in Experiment 1 completed version B. The standard time limit of 45 minutes was used here. The number of correct solutions was used as the dependent measure.

Spatial Locations and Relations Task: Participants also completed a novel WM task that required participants to either maintain spatial relations or spatial locations (Figure 1). This task was modeled after the tasks used in previous studies (Ackerman & Courtney, 2012; Blacker & Courtney, 2016; Ikkai et al., 2014).

All stimuli were presented on a 50% gray background. A trial began with a 500ms fixation cross, presented in the middle of the screen. Next, a 500ms verbal cue indicated whether the trial would be a "Relation" or a "Location" trial. A sample array was then presented for 500ms, which contained 2 or 3 colored circles (each subtending 0.67°×0.49° of visual angle). The color of each circle was chosen randomly without replacement from red, green, yellow and blue. The circles in an array were presented between 2.0-2.9° of visual angle apart both horizontally and vertically, but within the same quadrant of visual space. After a 2000ms delay period, a test array was displayed for 500ms. Participants had a maximum of 1500ms to enter their response, which meant that the 1000ms following the offset of the test array was also considered the response period. Finally, a 300ms feedback display was presented where the fixation cross turned green for a correct response, red for an incorrect response, and blue if the response was slower than 1500ms.

For both trial types, there was a low load (i.e., sample array contained 2 colored circles) and a high load (i.e., sample array contained 3 colored circles) condition.

For Location trials (Figure 1), under low load, participants were instructed to imagine a line segment connecting the two sample circles and maintain the location of that line in memory across the delay period. These instructions were used to encourage participants to encode the exact spatial coordinates of one concrete object (i.e., the imaginary line segment), while the same number of circles (i.e., two) was presented on the screen for both Location and Relation trials. At test, participants were asked to decide whether or not the black test circle fell on that imaginary line segment. For match trials, the black test circle fell in the exact center between the two previously presented sample circles (i.e., in the center of the remembered imaginary line segment). For non-match trials, the black test circle fell between 1.6×2.2° and 2.1×2.9° of visual angle from the position of that center point in any direction between the sample circles.

For Location trials, under high load, participants were instructed to remember the absolute locations of the three sample circles. At test, participants were asked to decide whether or not the black test circle was in one of the three sample locations or in a completely new location. For match trials, the black test circle fell in the identical location as one of the

sample circles and for non-match trials it fell between $1.6 \times 2.2°$ and $2.1 \times 2.9°$ of visual angle from any of the three sample locations.

While the low and high load Location trials consisted of different instructions to the participant, the key factor is that under low load participants were asked to maintain one spatial location (i.e., the location of the imaginary line segment) and under high load participants were asked to maintain three spatial locations (i.e., the locations of the three sample circles). The imaginary line manipulation in the low load trials was necessary to equate the number of sample circles displayed for Location and Relation trials. Further, previous work using this task has shown that the load manipulation for Location trials activates brain regions that are typically found with load increases in other types of visuospatial WM (Blacker & Courtney, 2016), such as posterior parietal cortex (e.g., Todd & Marois, 2004) and frontal eye fields (e.g., Leung, Seelig, & Gore, 2004).

For Relation trials (Figure 1), under low load, participants were instructed to encode and maintain the relative vertical positions of the two sample circles (e.g., red is above blue). Upon test, participants indicated whether or not the circles in the test array had the same relative vertical positions as the sample circles. For Relation trials, under high load, participants were instructed to encode and maintain the three possible vertical relationships between the sample circles (e.g., green is above yellow, yellow is above red, red is below green). As with low load, at test, participants indicated whether or not the circles in the test array had the same relative vertical positions as the sample circles. Note, which of the three relationships was tested was unpredictable, which forced participants to maintain all three relationships during the delay period. The horizontal relationship was never task-relevant in Relation trials.

There are a few crucial aspects of the task design worth elaborating on. First, regardless of trial type, under low load, participants were asked to encode and maintain one piece of information: either one spatial location (Location trials) or one spatial relation (Relation trials) and under high load, participants were asked to encode and maintain three pieces of information: either three spatial locations or three spatial relations. Second, trial type was pseudorandomly presented so participants could not predict what trial type they would see until the cue. Load was uncued, so the participants were unaware of the load until the sample array appeared. Third, the sample array circles were always presented in one quadrant of the display and the test array circles were always presented in the same quadrant as the sample circles, although in the Relation task the test circles were never in the exact same locations as the sample circles. For all trial types, participants pressed one button for a "match" response and another for a "non-match" response and these response key mappings were counterbalanced across participants. Participants completed a total of 256 trials.

Data Analysis—The goal of this initial experiment was to investigate individual differences in WM for spatial relations versus locations and how those differences predicted n-back, complex span, and *Gf* performance. To do so, we tested partial correlations between Relation WM accuracy, while controlling for Location WM accuracy, and n-back, Symmetry Span, and BOMAT. We also tested for partial correlations between Location WM accuracy, while controlling for Relation WM accuracy, and n-back, Symmetry Span, and BOMAT. For

the Location and Relation task we focused our analysis on high load trials, because previous work has shown that individual differences in WM performance are more evident when load is high (Cusack, Lehmann, Veldsman, & Mitchell, 2009; Linke, Vincente-Grabovetsy, Mitchell, & Cusack, 2011). However, we report parallel analyses for the low load condition in the Supplementary Online Material.

Results

One participant had below chance level accuracy (<50%) on the Locations/Relations task and was not included in any analyses. To report complete results for the Locations/Relations task, we tested a 2 (trial type: Location vs. Relation) × 2 (load: low vs. high) repeated-measures ANOVA. A main effect of load emerged, $F(1,23)=66.61$, $p<0.001$, partial $\eta^2=0.74$, with accuracy being higher for low load compared to high load. However, neither the main effect of trial type, $F(1,23)=1.04$, $p=0.32$, nor the trial type × load interaction, $F(1,23)=1.72$, $p=0.20$, reached significance. Table 1 includes descriptive statistics for each level of the task.

Next, we examined whether Relation WM accuracy was correlated with average d' on the spatial n-back task, while controlling for Location WM accuracy. A significant correlation emerged, $R(21)=0.59$, $p=0.003$, whereby higher Relation WM accuracy was associated with higher n-back d' values. Conversely, Location WM accuracy was not significantly correlated with n-back performance, when controlling for Relation WM accuracy, $R(21)=0.30$, $p=0.16$. Next, we examined how these two types of WM related to Symmetry Span performance. Neither Relation WM nor Location WM accuracy, while controlling for the other, was significantly correlated with Symmetry Span Score, ps 0.064. Moreover, Relation WM accuracy was significantly correlated with n-back d' even after controlling for Symmetry Span Score, $R(21)=0.49$, $p=0.018$, which demonstrates that n-back performance, above and beyond Symmetry Span performance, is related to Relation WM. Finally, we examined correlations with BOMAT score. Relation WM accuracy, controlling for Location accuracy, was significantly correlated with BOMAT score, $R(21)=0.44$, $p=0.034$, whereby better Relation WM performance was associated with better BOMAT performance. Conversely, Location WM accuracy was not significantly correlated with BOMAT score, when controlling for Relation WM accuracy, $R(21)=0.16$, $p=0.47$. These results are illustrated in Figure 2. Here it was possible to test whether these above partial correlation coefficients were significantly different from one another for the BOMAT analyses. The difference here indeed did not reach significance, $p=0.33$ (two-tailed). However, our goal was to determine if Relation and Location WM predict BOMAT performance while controlling for the other type of WM. We have used this approach elsewhere (Blacker, Weisberg, Newcombe, & Courtney, 2017) and argue that these partial correlations are evidence for a stronger relationship between Relation WM and BOMAT performance compared to Location WM and the BOMAT.

Given that Relation WM was significantly correlated with n-back and BOMAT performance, it would be expected that n-back and BOMAT performance would also be highly correlated. In line with previous work demonstrating the robust relationship between WM and *Gf* performance (e.g., Engle et al., 1999), here we did find a significant positive correlation between BOMAT score and n-back d', $R=0.58$, $p=0.003$.

Discussion

The results of Experiment 1 represent a proof of concept that performance on a relational WM task is predictive of performance on both an n-back task and a measure of *Gf*. This supports our initial hypothesis that the n-back task relies more heavily on relational processing than complex span tasks do. It also demonstrates that a measure of *Gf*, the BOMAT, is related to relational WM performance. Therefore, we next examined whether this reliance on relational WM could explain differences in transfer between these two tasks in a training study, as well as whether n-back and complex span training would result in significantly different changes in alpha power following training. If n-back training were to have a greater effect than complex span training on an EEG measure related to relational WM, then it would provide support for the idea that the involvement of relational WM might underlie its greater effectiveness for cognitive training and transfer.

Experiment 2: Training Study

Method

Participants—136 neurologically healthy adults (age: *M*=21.18, *SD*=3.12; 44 males) participated for monetary compensation. All participants were recruited through flyers and online university announcements. Recruitment materials stated "Participants needed for research on cognitive training." No participants from Experiment 1 participated in the training study. Participants who completed the study received an average of $215 (*SD*=11)[2]. All participants had normal or corrected-to-normal vision, and gave written informed consent approved by the Institutional Review Boards of both Johns Hopkins University and the Johns Hopkins Medical Institutions.

General Study Procedures—The general study structure is illustrated in Figure 3. Upon enrollment in the study, participants were randomly assigned to a training group as noted above and then completed a 2.5hr pre-training assessment session. This initial session consisted of the informed consent, demographic questionnaire completion, training program installation onto their personal laptop, completion of 6 assessment tasks, instructions and practice trials for the EEG task, and instructions and practice trials for their respective training task. All of these tasks are described in detail below. The second session was a 2hr pre-training EEG session, which included completion of several questionnaires and a separate WM task while EEG was recorded. After the pre-training EEG session, participants were instructed to begin their at-home training sessions immediately and to complete 5 sessions per week for 4 weeks. Upon completion of the 20 training sessions, participants completed a 2hr post-training EEG session where they completed the same WM task while EEG was recorded. Finally, participants completed a 1.5hr post-training assessment session where they completed the 6 assessment tasks again and completed a study exit survey.

Training Groups and Tasks—Participants were randomly assigned to one of three training groups upon enrolling in the study: dual n-back training (DNBT), symmetry span

---

[2]Variation in payment arose from participants receiving a completion bonus of $15 upon completing all study sessions *and* maintaining above chance level performance during training (e.g., two consecutive sessions of average performance below chance for each training task resulted in the participant not receiving the bonus). Participants were aware of the contingencies of this bonus.

training (SST) or permuted rule operations (PRO). During the initial study session, we installed MATLAB (The MathWorks, Natick, MA) and Psychtoolbox (Brainard, 1997; Pelli, 1997) onto each participant's personal laptop, along with the scripts for their assigned training program. Each training session was set to last 30 minutes. Participants emailed their data files to the laboratory after every session.

Dual N-back Training (DNBT): Similar to Jaeggi et al. (2008), one group of participants performed a dual auditory and visuospatial *n*-back task simultaneously (Figure 4A). Participants saw a blue square on a gray background in one of 8 spatial locations on a computer screen and at the same time heard one of 8 letter sounds (i.e., consonants). These stimulus pairs were presented every 3 seconds. Participants were asked to judge whether each stimulus was in the same location or the same letter as the stimuli *n* before. Participants responded by pressing: the "f" key if the location was the same as the location n-back, the "j" key if the letter was the same as the letter n-back, and both "f" and "j" if both the location and letter were the same as n-back. Participants began training at the 1-back level. A sequence of 20+*n* stimuli was presented in a given block, where there were 4 location-only targets, 4 letter-only targets and 2 dual targets in each block. The training was adaptive, whereby participants were given feedback on their accuracy at the end of each block. If accuracy was >85%, the next level would increase by 1 (i.e., a 2-back), whereas if accuracy was <70%, the next level would decrease by 1 (or stay the same if they were on a 1-back), and in all other instances the level would remain unchanged[3]. While Experiment 1 provided evidence of a relationship between single n-back performance and *Gf* and Relation WM, here we used a dual n-back task for training because it is more common in the literature to use dual n-back tasks for training purposes and there is evidence that dual and single n-back task performance are roughly equally correlated with performance on measures of *Gf* (Jaeggi et al., 2010).

Symmetry Span Training (SST): An adaptive version of the automated symmetry span task (Kane et al., 2004) was used here, similar to that used by (Richmond et al., 2011). Participants recalled sequences of red-square locations within a matrix against a background symmetry judgment task (Figure 4B). The participant's task was to remember the order and locations of red squares presented in a 4×4 grid. Each red square was presented for 650ms. Interleaved between each red square, a white 8×8 matrix was presented, with some squares filled in black and participants were asked to decide if the image was symmetrical along its vertical axis. A total of 96 symmetry images were utilized (50% symmetrical, 50% asymmetrical). Participants responded by clicking a box labeled "yes" or "no" with the mouse. After all red squares and symmetry judgments were presented, a blank 4×4 grid was presented and the participant recalled the red square locations in the temporal order they were presented by clicking in the blank grid with the mouse. This training task was made

---

[3]Our adaptivity criteria for the DNBT is a departure from that used by Jaeggi et al., (2008) and many subsequent studies using this paradigm for training. In Jaeggi et al., (2008), with 12 targets per block (same as used here), if participants made fewer than three mistakes per modality they advanced to the next *n*-level. Thus, accuracy was considered separately for each modality. Here, we considered accuracy across both modalities for our cut-off values. In the present study, participants had to perform above 85% on both modalities considered together to move up an *n*-level. This more stringent criteria likely explains why our participants on average did not advance as far up the *n*-levels as some previous studies (see Figure 8) as it would have been more difficult to advance up a level and easier to fall back down a level with this criteria.

adaptive in two ways. Training began with 2 memory items and 1 symmetry judgment following each memory item. The next level would contain 2 memory items and 2 symmetry judgments following each memory item, then 3 symmetry judgments. After 3 symmetry judgments the number of memory items would increase to 3, but the symmetry judgments would be 1 following each memory item, then 3 memory items and 2 symmetry judgments and so on. If memory accuracy was >85% and symmetry accuracy was >75%, the next level would increase by 1, whereas if memory accuracy was <70% or symmetry accuracy was <60% the next level would decrease by 1 (or stay the same if they were on level 1), and in all other instances the level would remain unchanged.

Permuted Rule Operations (PRO): The PRO training group served as an active control group as the training did not involve a WM task, but was a cognitively challenging and adaptive training program. The task was adapted from the task used by Cole, Bagic, Kass, and Schneider (2010). Participants were presented a set of 3 rules and then asked to apply those rules to two word stimuli as quickly as possible (Figure 4C). Four semantic rules, four decision rules, and four response rules were used. The semantic rule consisted of sensory semantic decisions (i.e., "is it sweet?"). The four possible semantic rules were: SWEET, GREEN, SOFT and LOUD. The decision rule specified (using logical relations) how to respond based on the semantic decision outcome(s) for each trial. The SAME rule required that the semantic answer was the same ("yes" and "yes" or "no" and "no") for both words, the DIFFERENT rule required that the answer was different for the two words, the SECOND rule required that the answer was "yes" for the second word, while the NOT SECOND rule required that the answer was "no" for the second word. The motor response rules specified what button to press based on the decision outcome. The task instructions made explicit reference to the correct response for a "true" outcome, while participants knew (from the practice session) to use the other finger on the same hand for a "false" outcome. For all participants the "f" and "d" keys corresponded to left index and middle finger responses, respectively; whereas the "j" and "k" keys corresponded to the right index and middle finger responses, respectively. As shown in Figure 4C as an example trial, if the participant was presented with the rules: SWEET, SAME, LEFT MIDDLE and the words were CHOCOLATE and GRASS then the participant would press the "f" key (i.e., the left index key) to indicate that the outcome was false (i.e., chocolate and grass are *not* the same on the attribute of sweetness).

As with DNBT and SST, the PRO training was adaptive. Participants started with only one combination of rules and with a self-paced timing. Self-paced meant that participants could spend as much time as they needed encoding the rules, then they initiated the trial by pressing the space bar, which would bring up the two words they had to apply the rules to. In the beginning, participants had 1500ms to respond. The next level would include a time-limited encoding duration of 1500ms. As participants progressed through the levels the encoding time was shortened to 500ms and then eventually to 300ms. Once participants were able to perform the task with the 500ms encoding duration, the number of rule combinations would increase. As all 64 of the rule combinations were mixed in, then the time in which the participant had to respond was shortened from 1500ms to 1000ms to 750ms to 500ms. These different manipulations all served to increase the difficulty of the

task as participants progressed through the various levels. Participants performed 32 trials per block. If accuracy was >85% in a given block, the next level would increase by 1, whereas if accuracy was <60% the next level would decrease by 1 (or stay the same if they were on level 1), and in all other instances the level would remain unchanged.

Attrition: A total of 73 participants completed the study. Sixty-three participants failed to complete the study. Nine participants did not have a computer that would run the study-related software required for the at-home training sessions, which classified these participants as failures in pre-enrollment screening. Additional reasons for dropping out included a change in exclusion criteria status, loss of interest, being too busy with schoolwork, inability to perform above chance on the Relation/Location WM task, or simply ceasing to answer emails from the experimenters. A detailed breakdown of these reasons and the number of corresponding participants is listed in Table S1 in the Supplemental Online Material. Participants were equally likely to drop out from each of the 3 training groups (see Figure 3). Considering individuals who dropped out either without contact or reason and those that did provide a reason for dropping out, we had an unusually high drop-out rate of 46 participants (34% of the initial sample). Given the large proportion of participants who did not provide us with a rationale for their withdrawal from the study, we can only speculate about the high attrition rate seen here. Our participants were students from the Johns Hopkins community and this population has a reputation for being very focused on coursework and academic outcomes. Our speculation is that many students opted to drop-out of our time-intensive study in order to prioritize coursework.

Assessment Tasks and Procedures—Experimental stimuli were controlled by MATLAB (The MathWorks, Natick, MA) using Psychophysics Toolbox extensions (Brainard, 1997; Pelli, 1997), and displayed on a 21" Apple iMac monitor. Participants were seated approximately 60cm from the monitor.

Object n-back: Participants were shown a sequence of object stimuli and they had to respond each time the current stimulus was identical in shape to the one presented *n* positions back in the sequence (Figure 5A). The stimuli consisted of 10 random shapes (Vanderplas & Garvin, 1959), which have been used in an n-back task previously (Jaeggi et al., 2003). The shapes were presented in one of four colors and participants were told that color was irrelevant. The timing, n-levels used, and dependent measure were identical to that reported in Experiment 1 for the spatial n-back.

Operation Span (Ospan): We used the Ospan task as a complex span measure of WM (Kane et al., 2004). As shown in Figure 5B, the task requires participants to recall a sequence of letter stimuli in the correct order in addition to completing a distracting processing task. We presented three sets of stimuli per set size (i.e., number of stimuli to be recalled) and the set sizes ranged from 4 to 8. The processing task involved quickly responding to whether a math problem was accurate or not (e.g., 4+7=11). The score was measured as the sum of letters recalled correctly in the correct order, regardless of whether the entire trial was recalled correctly.

Task Switching: Participants completed a task switching paradigm (Figure 5C) similar to that developed by (Rogers & Monsell, 1995). We chose to include this task in our assessment battery because we predicted that it would represent a near transfer measure for the PRO training group, which trained on a rapid rule learning and switching paradigm. A trial began with a 1000ms fixation period followed by a cue presented for 200ms. The cue either read "LETTER" or "NUMBER". The cue was followed by a 226ms delay period and then an alphanumeric pair of stimuli centrally presented. If cued LETTER, participants had to respond whether the letter was a vowel or consonant (e.g., left button for vowel, right button for consonant). If cued NUMBER, participants had to respond whether the number was even or odd (e.g., left button for even, right button for odd). Response mapping was counterbalanced across participants. Participants were instructed to respond as quickly as possible without losing accuracy. After a response, a 950ms intertrial interval (ITI) was presented. Response time (RT) was the dependent measure and was considered separately for repeat trials (i.e., trials where the previous trial was the same cue) and switch trials (i.e., trials where the previous trial was the other cue). Participants completed a total of 240 where 60% of trials were repeat trials. The critical dependent measure was switch cost (switch RT – repeat RT).

Bomat: Before and after training we assessed *Gf* using the BOMAT, as described in Experiment 1. Here, all participants completed Form A before training and Form B after training. A 20 minute time limit was enforced here, which has been used previously in training studies (Jaeggi et al., 2008). The time restriction was used to avoid ceiling level performance and to keep the total testing time as short as possible for all assessments.

Additional Assessment Tasks: Participants also completed two additional tasks before and after training: a change detection task (Luck & Vogel, 1997), and the Attentional Network Test (Fan, McCandliss, Sommer, Raz, & Posner, 2002). The task procedures and results are reported in the Supplemental Online Material. While the focus here was on near transfer and far transfer to *Gf* only, we report all measures in an effort to adhere to the methodological suggestions of Boot, Blakely, and Simons (2011) for cognitive training studies.

Motivation & Engagement Measures: Participants provided self-report ratings of motivation and engagement for each task and also completed an exit survey at the end of the study, which assessed how well they liked training and whether they thought they improved on the training task and/or the assessment tasks. Additional description and results of these measures can be found in the Supplemental Online Material.

EEG Task and Procedures—Experimental stimuli were controlled by MATLAB (The MathWorks, Natick, MA) using Psychophysics Toolbox extensions (Brainard, 1997; Pelli, 1997), and displayed on a 19" Dell LCD monitor. Participants were seated 92cm away from the monitor, and given a Logitech game controller to enter responses.

Spatial Locations and Relations Task: This task was identical to the one described above in Experiment 1 with the following exceptions. The fixation cross here appeared for 900ms and the feedback display appeared for 100ms. Also, each circle subtended 0.42°×0.31° of visual angle. Memory array circles were presented between 1.5-1.8° of visual angle apart

both horizontally and vertically, but again within the same quadrant of visual space. For non-match trials, the black test circle in Location trials fell between 1.1×1.5° and 1.4×1.8° of visual angle from the position of that center point in any direction between the sample circles. Finally, participants completed 88 practice trials during their initial pre-assessment visit without EEG being recorded in order to learn the task. Then during the EEG session, participants completed a total 512 trials.

EEG Data Acquisition & Analysis

EEG Recording: EEG data were recorded at 47 sites covering the whole scalp with approximately uniform density using an elastic electrode cap referenced to the average of all electrodes during recording (WaveGuard cap with 47-channel Duke, equidistant electrode placement, layout: Advanced Neuro Technology, The Netherlands). Electrode impedance was kept below 15kΩ. All EEG electrodes were recorded continuously in DC mode at a sampling rate of 512Hz using an anti-aliasing filter with a 138Hz cutoff and a high-impedance ANT WaveGuard amplifier.

Preprocessing: Data were analyzed using the Fieldtrip software package (Oostenveld, Fries, Maris, & Schoffelen, 2011). Data were first high-pass filtered at 0.5Hz, and then segmented into epochs covering the time from 2.0sec before to 5.5sec after the onset of the fixation display in each trial. Independent components analysis (ICA) was performed on the epoched data, and the eye blink component was identified and removed for each participant's data. After eye blink correction, EEG waveforms from frontal electrodes (i.e., RE1/LE1) were visually inspected to identify voltage fluctuations (i.e., fluctuations greater than 18.75μV or less than -18.75μV) typical of eye movements or gross motor movements. Ocular artifacts are characterized by high, >15μV amplitude, step-like potentials of opposite polarity. Here we used 18.75μV as the cut-off to be slightly more conservative, as well as to be consistent with our groups' previous EEG work with variants of this task. Our group has empirically learned that this is the most sensitive threshold for eye-blink detection, based on multiple datasets using this electrode configuration and equipment. Trials containing these movement artifacts were rejected entirely. EEG data were analyzed only for correct trials.

To maintain sufficient statistical power for each session, participants with more than 40% trial rejection due to any combination of artifacts or behavioral errors were excluded (N=11). Both pre- and post-training EEG data were excluded for those 11 participants. One additional participant's pre-training EEG data was excluded because they failed to complete the post-training EEG session. The final sample of analyzed EEG data consisted of 2 sessions for 61 participants (DNBT: N=20, SST: N=22, and PRO: N=19).

Spectral analysis: Power spectra were calculated using a time-frequency transformation based on multiplication in the frequency domain from 1 to 30Hz with 0.5Hz increments using a hanning taper applied to short sliding time windows (Percival & Walden, 1993) every 100ms. An adaptive time window of five cycles for each frequency ($T = 5/f$) was applied.

Statistical Analysis: To obtain statistics corrected for multiple comparisons we used nonparametric permutation tests (Maris & Oostenveld, 2007; Nichols & Holmes, 2002). This procedure controls for Type I error by calculating the cluster-level statistics by randomizing trial labels at each iteration. First, spectral data from each of the 47 electrodes across the scalp were averaged over the time and frequency range of interest. Our frequency range of interest was alpha (8-13Hz) based on previous work described above in the Introduction. Our time range of interest was the WM delay period (i.e., 1.9 – 3.9 seconds after the onset of fixation), but we excluded the first 500ms of the delay period because this time period likely contained sensory-evoked response activity from the memory array (e.g., van Gerven et al. 2009; for a more detailed discussion of this topic, also see Bastiaansen, Mazaheri, & Jensen, 2012). This timeframe left 1500ms of delay period activity to analyze and given the need to have a minimum of ~10 cycles to achieve a reliable measure of power, we could not break this time period down any further when examining 8-13Hz activity. Next, a t-value was calculated at each electrode. For each iteration randomizing trial labels, clusters of electrodes where the alpha-level was <0.05 were identified, and their t-values were summed. The largest sum of t-values was used as a t-statistic. This procedure was repeated 5000 times to create the null distribution. The p-value was estimated according to the proportion of the null distributions exceeding the observed cluster-level t-statistic. It is convention with this analysis approach to report only the p-value for each cluster of electrodes.

Here our main analysis of interest was a group (DNBT, SST, PRO) × session (pre-, post-training) interaction. This was done by first calculating the post – pre difference and then testing for differences between pairs of groups (i.e., DNBT vs. SST, DNBT vs. PRO, SST vs. PRO). Fieldtrip does not allow for a 3-level interaction permutation test, therefore we imposed a Bonferroni multiple comparison correction for the three comparisons (i.e., interaction tests with a p-value less than 0.017 were considered statistically significant). Significant group × session interactions were followed up with post-hoc contrasts by group and session.

Results

Behavioral Results—For each assessment task, we were interested in whether any group improved significantly more after training compared to before training compared to the other two groups. Therefore, we tested 3 (group: DNBT, SST, PRO) × 2 (session: pre, post) repeated-measures ANOVAs on the dependent measure for each assessment task. Our primary assessment measures were the object n-back task, Ospan, task switching, and the BOMAT and reliability for each of these measures can be found in Table 2.

Object n-back: For the object n-back task, our dependent measure was the average d' score across all 5 n-back levels. Any participant with performance >2SD below the mean for either session was excluded from any analyses (N=4). These exclusions resulted in a total of 69 participants remaining in the analysis. A main effect of session emerged, $F(1,66)=21.64$, $p<0.001$, partial $\eta^2=0.25$, where participants improved significantly from pre- to post-training. The main effect of group also reached significance, $F(2,66)=3.87$, $p=0.026$, partial $\eta^2=0.11$, with the DNBT showing the highest performance followed by SST and PRO,

respectively. Critically, the group × session interaction was significant, $F(2,66)=5.28$, $p=0.007$, partial $\eta^2=0.14$. Figure 6A shows the performance gains for all 3 training groups illustrating the largest gains for the DNBT group.

Planned contrasts demonstrated that the DNBT group showed significantly higher performance after training compared to before training, $t(22)=5.82$, $p<0.001$, as did the SST group, $t(23)=2.29$, $p=0.031$; whereas the PRO group's performance did not differ from pre- to post-training, $t(21)=0.50$, $p=0.62$. Moreover, the DNBT group showed significantly more improvement from pre- to post-training than both the SST group, $t(45)=2.02$, $p=0.049$, and the PRO group, $t(43)=3.34$, $p=0.002$. The SST group did not improve more so than the PRO group, $t(44)=1.24$, $p=0.22$. Together, these results reveal significant near transfer for the DNBT group to the object n-back task.

Ospan: The dependent measure was the Ospan score, which is the sum of correctly recalled letters in the correct order. Two participants did not complete the pre-training Ospan task. Following previous work, we excluded any participant with accuracy on the processing task <75% for either session (N=13). The high rate of participants with poor accuracy was driven by several participants not understanding the task during the pre-training session due to experimenter error. Further, any participant with memory performance >2SD below the mean for either session was also excluded from any analyses (N=3). These exclusions resulted in a total of 55 participants remaining in the analysis. Processing task accuracy for the remaining 55 participants was 93.4% (SD=4.9). A main effect of session emerged, $F(1,52)=29.10$, $p<0.001$, partial $\eta^2=0.36$, where participants improved significantly from pre- to post-training. The main effect of group was not significant, $F(2,52)=0.47$, $p=0.63$. The group × session interaction did not reach significance, $F(2,52)=0.75$, $p=0.48$. Figure 6B shows the performance gains for all 3 training groups, illustrating that there was no significant difference in pre-training performance between the three groups and that the two WM training groups, SST and DNBT, did show numerically greater gains compared to the PRO group.

Task Switching: The dependent measure for Task Switching was the switch cost RT, i.e., switch trial RT - repeat trial RT, which means that greater values correspond to a larger cost in RT for switching rules. RT was only analyzed for correct trials. Three participants did not complete the pre-training Task Switching measure. Participants with task accuracy <75% for either session were excluded (N=5) or with mean RT >2SD above the mean was also excluded (N=1). These exclusions resulted in a total of 64 participants remaining in the analysis. Task accuracy for the remaining 64 participants was 95.5% (SD=3.6). A main effect of session emerged, $F(1,61)=5.90$, $p=0.018$, partial $\eta^2=0.09$, where participants improved significantly from pre- to post-training, meaning they had smaller switch costs after training compared to before training. The main effect of group was not significant, $F(2,61)=0.64$, $p=0.53$. Critically, the group × session interaction was significant, $F(2,61)=3.23$, $p=0.046$, partial $\eta^2=0.10$. Figure 6C shows the performance gains for all 3 training groups, illustrating the largest gains (i.e., a decrease in switch cost) for the PRO group.

Planned contrasts demonstrated that the PRO group showed significantly smaller switch costs after training compared to before training, $t(19)=2.85$, $p=0.01$; whereas neither the DNBT group's performance, $t(20)= -0.34$, $p=0.74$, nor the SST group's performance, $t(22)=1.44$, $p=0.16$, differed from pre- to post-training. Moreover, the PRO group showed significantly greater improvement from pre- to post-training than the DNBT group, $t(39)=2.40$, $p=0.021$, and a trend toward greater improvement compared to the SST group, $t(41)=1.46$, $p=0.1$. The DNBT group did not improve more so than the SST group, $t(42)=1.21$, $p=0.23$. Together, these results reveal significant near transfer for the PRO group to the task switching paradigm.

While switch cost was the primary dependent measure of interest for the task switching task, we also tested a 3 (group: DNBT, SST, PRO) × 2 (session: pre, post) repeated-measures ANOVA on task accuracy to ensure the RT results could not be explained by a difference in accuracy or a speed-accuracy tradeoff. A main effect of session emerged, $F(1,61)=11.77$, $p<0.001$, partial $\eta^2=0.16$, with accuracy being higher after training compared to before training (pre-training $M=96\%$, post-training $M=94\%$). Importantly, neither the main effect of group, $F(2,61)=0.95$, $p=0.39$, nor the group × session interaction, $F(2,61)=0.29$, $p=0.75$, approached significance, which suggests that the selective improvement of the PRO group on switch cost following training cannot be explained by a speed-accuracy tradeoff.

<u>Bomat:</u> The dependent measure was the total number of correctly solved problems within the time limit. No outliers were present on this measure (i.e., no score was $\pm 2SD$ from the group mean). The main effect of session was significant, $F(1,70)=28.74$, $p<0.001$, partial $\eta^2=0.29$, with performance being higher after training compared to before. Neither the main effect of group, $F(2,70)=0.09$, $p=0.91$, nor the group × session interaction, $F(2,70)=0.77$, $p=0.47$, approached significance. Figure 7 illustrates the group means by session.

<u>Predictors of Near Transfer:</u> While previous studies have shown near transfer, it has also been demonstrated that there are individual differences in performance on training and near transfer (Jaeggi et al., 2014). Figure 8 illustrates group average session by session performance on the three training tasks. First, we were interested in whether performance on training was correlated with the degree of near transfer. To standardize each training groups' performance, we calculated the maximum training level reached throughout the 20 sessions and then z-scored that value for each group. A correlation analysis (not corrected for multiple comparisons) showed that maximum training level showed a trend toward being positively correlated with improvement on the object n-back task, for the DNBT group, $R(47)=0.29$, $p=0.05$, but not for the SST group, $R(24)=0.24$, $p=0.25$, or PRO group, $R(22)=0.07$, $p=0.76$. Figure 9A shows that better performance during training for the WM groups was associated with greater gains on the object n-back task. Further, we found that baseline performance (i.e., pre-training) on the object n-back task was related to improvement on that task following training, for our WM groups, but not the PRO group. To examine this, we performed a median split on pre-training performance on the object n-back task and then compared gain scores (post – pre) within each group (see Figure 9B). For the DNBT and SST groups, the low WM participants saw greater gains after training compared to high WM participants, DNBT: $t(21)=3.01$, $p=0.007$, SST: $t(22)=3.51$, $p=0.002$, but this

difference was not significant for the PRO group, $t(20)=1.52$, $p=0.14$, showing that this effect was specific to individuals who trained on a WM task.

Behavioral Results during EEG—Similar to the other transfer tasks described above, we were interested in whether any group improved significantly more after training compared to the other two groups on the Locations/Relations WM task performed during EEG recording. We analyzed the behavioral data from the 61 participants who completed the study and had sufficient EEG data for analysis (see details above in the Methods)[4]. First, we tested a 3 (group: DNBT, SST, PRO) × 2 (session: pre, post) × 2 (load: low, high) × 2 (trial type: Location, Relation) repeated-measures ANOVA on accuracy. Significant main effects of load, $F(1,58)=639.07$, $p<0.001$, partial $\eta^2=0.92$, and trial type emerged, $F(1,58)=136.05$, $p<0.001$, partial $\eta^2=0.70$, with accuracy being higher on low load and Relation trials, respectively. The session × trial type interaction approached significance, $F(1,58)=3.90$, $p=0.05$, partial $\eta^2=0.06$, with greater improvement on the Relation trials after training compared to the Location trials. Further, the load × trial type interaction was also significant, $F(1,58)=264.13$, $p<0.001$, partial $\eta^2=0.82$, with higher load decreasing accuracy to a greater degree for Location trials compared to Relation trials. All other main effects and interactions did not reach significance, all Fs 1.80, all ps 0.17. Given the lack of a significant interaction including the session and group factors, these results suggest no significant transfer to this WM task (see Figure 10).

EEG Results—To examine whether changes in neural activity occurred in response to training, we tested a group × session interaction on the delay period alpha (8-13Hz) power separately for Location and Relation trials. Here we only present analyses from high load trials, but group × session interaction results for low load trials are detailed in the Supplementary Online Material. For high load Location trials (see Figure 11A), there was a significant interaction in a group of frontal electrodes when comparing the DNBT and SST groups, $p=0.013$, and also when comparing the DNBT and PRO groups, $p=0.009$, which was in a similar group of frontal electrodes. However, the group × session interaction comparing SST and PRO groups was not significant, $p>0.05$. To follow up on the significant interactions, we examined spectral power differences within each testing session as well as within each group. Before training, there was no significant difference in alpha power between the DNBT group and either the SST or PRO groups. However, after training, the DNBT group showed significantly more alpha power in a group of frontal electrodes compared to the SST group, $p<0.05$, and in a cluster of frontal electrodes and a separate cluster of posterior electrodes when compared to the PRO group, $ps<0.05$. We also examined whether each group showed significant changes in alpha power by comparing the post- to pre-training alpha power. While the PRO and SST groups showed no significant change in alpha power, the DNBT showed significantly more alpha power after training compared to before training, $p<0.005$, in a group of frontal electrodes.

For high load Relation trials (see Figure 11B), there was a significant interaction in a group of left frontal electrodes when comparing the DNBT and PRO groups, $p=0.017$, but not

---

[4]The direction of the effects and significance remain unchanged when the analyses are tested on all 72 participants that completed both pre- and post-training EEG sessions.

when comparing the DNBT and SST groups, $p>0.05$. Also, the group × session interaction comparing SST and PRO groups was not significant, $p>0.05$. To follow up on the significant interaction, we examined spectral power differences within each testing session as well as within each group. Before training, there was no significant difference in alpha power between the DNBT group and either the SST or PRO groups. However, after training, the DNBT group showed significantly more alpha power in a group of right frontal electrodes compared to the SST group, $p<0.05$, and in a cluster of frontal electrodes and a separate cluster of posterior electrodes when compared to the PRO group, $ps<0.05$. We also examined whether each group showed significant changes in alpha power from to pre-training to post-training. While the PRO and SST groups showed no significant change in alpha power, the DNBT group showed significantly more alpha power after training compared to before training, $p<0.05$, in a group of left frontal electrodes.

Baseline Correlations—We found no evidence in our training data that DNBT has a differential effect on relational WM. However, in Experiment 2 we have a large baseline data set that could be used to attempt to replicate the initial correlational findings from Experiment 1. We therefore tested the same partial correlations as in Experiment 1, here with our baseline sample of 119 participants who completed both pre-training sessions (i.e., assessments and EEG). The partial correlation between Relation WM and pre-training object n-back performance while controlling for Location WM did not reach significance, $p=0.29$. Likewise, there was no relationship between Location WM and n-back performance when controlling for Relation WM, $p=0.81$. Similarly, there was no significant relationship between Relation WM and OSpan or BOMAT when controlling for Location WM, $p=0.16$ and $p=0.64$, respectively. There was also no significant relationships between Location WM and OSpan or BOMAT when controlling for Relation WM, $p=0.63$ and $p=0.61$, respectively.

Discussion

Experiment 2 demonstrated clear near transfer to the object n-back task for both of our WM training groups, as well as robust near transfer to a task switching paradigm for our active control group. Moreover, near transfer for our WM training groups was predicted by baseline WM level, whereby individuals with lower baseline WM improved more after training compared to their high WM counterparts. While each group improved on the BOMAT, we found no specific evidence of far transfer to *Gf* for any one training group. In fact, the PRO group showed the largest numerical gains on the BOMAT of the three groups. Finally, we found that the DNBT group showed significant increases in frontal alpha power compared to the other two training groups. While we did not find support for our hypothesis of specific improvement in relational WM driving the DNBT group's transfer, these results are consistent with the idea that DNBT may be a more effective training task as compared to complex span tasks. Further implications are discussed below.

General Discussion

We initially hypothesized that a critical difference between n-back and complex span WM tasks is their reliance on the ability to extract and maintain relational information in WM. We predicted that this difference might account for differences in training efficacy seen in the literature, whereby n-back tasks seem to produce more effective far transfer. In

Experiment 1, we found correlational evidence in support of this hypothesis whereby performance on a relational WM task predicted performance on an n-back task and a measure of *Gf*, but not a complex span task. We followed up on this finding with a training study in Experiment 2, which is the first study to directly compare n-back and complex span WM training in the same study. While we found strong near transfer effects, we saw no evidence of far transfer to *Gf*. We did not find evidence to support the notion that the effectiveness of DNBT is due to a dependence on relational WM. Further, we failed to replicate our findings from Experiment 1 of a relationship between relational WM and n-back performance in the larger sample in Experiment 2. While we ultimately did find support for our initial hypothesis about the role of relational WM in n-back training, our results demonstrated that DNBT did elicit robust near transfer and significantly greater neural changes compared to the other two training groups.

Our finding of near transfer but not far transfer, even with the DNBT group, is perhaps not surprising. There have been inconsistent results in the literature surrounding far transfer, but there have been far fewer inconsistencies with respect to near transfer, i.e., training on one WM task and finding evidence of improvement on a separate, untrained WM task. For example, training on one type of n-back task transfers to improvements on untrained versions of the n-back task (Colom et al., 2013; Jaeggi, Studer-Luethi, et al., 2010; Li et al., 2008). Similarly, training on one type of complex span WM task transfers to improvements on other complex span tasks (Chein & Morrison, 2010; Richmond et al., 2011). Consistent with this previous work, here we found that DNBT resulted in robust near transfer to an untrained n-back task. Moreover, the SST group also showed significant improvement on the object n-back task, but not significantly more so than the active control group. Importantly, we found clear near transfer from our active control group to a task switching paradigm, which suggests that this group was indeed an adequate control and that our WM training groups' transfer was specific to WM tasks. Surprisingly, we did not see significant near transfer from the SST or DNBT groups to the Ospan task even though the results were in the expected direction. There are at least two potential reasons for this null result. First, we had a particularly high number of outliers on the Ospan task as noted above in Experiment 2, which may have underpowered the critical group × session interaction. Second, SST and Ospan cross modality boundaries, with symmetry span being visuospatial and Ospan involving verbal, letter stimuli. This fits with previous evidence that intra- and across-modal training effects are separable (Schneiders, Opitz, Krick, & Mecklinger, 2011), but on the other hand there is evidence of cross-modality transfer (Buschkuehl et al., 2014; Jaeggi et al., 2014). It is possible that if we had used another visuospatial complex span task, we might have seen more robust near transfer; however, an important direction for future training studies will be to elucidate the role of stimulus modality in transfer.

Despite the near transfer pattern seen in Experiment 2, we found no evidence of far transfer to BOMAT performance. All 3 training groups showed significant improvement on the BOMAT after training compared to before training, but no one group showed significantly more improvement, which suggests that these increases may simply be practice effects. While we predicted far transfer *a priori*, our results are consistent with previous failures to find far transfer to *Gf* following WM training. Interestingly, two prominent examples of failure to find far transfer also failed to find any improvement from pre- to post-training on

measures of *Gf* following training (Redick et al., 2013; Thompson et al., 2013), which is unusual given the expectation of practice effects. Here we did observe significant improvement on the BOMAT from pre- to post-training, but that pattern was present for all three training groups. In light of this observation, one possible explanation for our results may be our choice of active control group. The PRO training involves rapid rule learning and the ability to quickly encode and switch between rules. It may be possible that these active ingredients helped participants to perform better on the BOMAT after training by increasing speed of processing and/or cognitive flexibility. This idea is speculative, but future work should address the effectiveness of PRO training in eliciting transfer effects. Furthermore, the choice of appropriate active control groups is an ongoing issue that will play a key role in moving the cognitive training field forward.

In Experiment 1, we found evidence that n-back performance was associated with relational WM performance supporting our initial hypothesis. However, in Experiment 2, neither the behavioral nor the EEG data demonstrated support for a specific role of relational WM in DNBT performance effects. And we also did not observe the same relationship between relational WM and n-back performance in our larger sample in Experiment 2. Here we sought to test out the notion that DNBT may exert different training effects than complex span because of a greater reliance on relational WM, as was initially suggested by the results of Experiment 1. This hypothesis was not supported, however, by the results of Experiment 2. It may be the case that DNBT does utilize relational WM, but additional factors also contribute to its training effectiveness, which diluted the transfer to the relation WM task. An alternative approach would be to specifically design a training paradigm that improves relational WM more directly. Future studies could test the effectiveness and transfer effects of such a training paradigm. Nonetheless, our EEG data did support the original motivation for our hypothesis, which was that DNBT seems to be more effective than other training approaches. Our results showed that only DNBT produced significant changes in alpha power, with significant increases in delay period frontal alpha power after training compared to before training for both Location and Relation trials. Neither the SST nor the PRO group showed any significant differences from pre-to post-training. Alpha activity over frontal brain regions have been interpreted as representing top-down modulation of WM (Zanto et al., 2011), which may be the mechanism by which DNBT is exerting its effectiveness; however, future studies will be needed to directly test this idea. Further, this lack of support for the role of relational WM in transfer, would benefit from future studies examining other possible factors that underlie WM training and transfer effectiveness. A recent example of such an effort provides evidence for the role distractor filtering in improving WM capacity via training (Li, He, Wang, Hu, & Guo, 2017).

Our findings with frontal alpha are consistent with the idea that training is influencing the frontoparietal network, which has also been shown in two previous studies using EEG. Jausovec and Jausovec (2012) found that WM training, on a diverse set of WM tasks, resulted in changes in alpha and theta event-related synchronization (ERS) when participants performed transfer tasks. However, in their study the EEG data were averaged across multiple transfer tasks that included measures of *Gf*, verbal analogy, and spatial rotation. This approach makes it difficult to pinpoint the changes in oscillatory synchronization to a specific aspect of training or transfer. Another study by Kundu, Sutterer, Emrich, and Postle

(2013) used EEG and TMS to demonstrate that transfer following WM training was supported by changes in task-related effective connectivity in frontoparietal and parieto-occipital networks that were engaged by the training and transfer tasks. While very few studies have examined changes in EEG measures due to training, there is a small body of work using fMRI that also suggests that training is associated with functional changes that have been localized to posterior parietal and prefrontal cortices (Beatty et al., 2015; Buschkuehl, Hernandez-Garcia, Jaeggi, Bernard, & Jonides, 2014; Salminen, Kuhn, Frensch, & Schubert, 2016; for an fNIRS study see, McKendrick et al. 2014; Schneiders et al., 2011; Thompson, Waskom, & Gabrieli, 2016; Vartanian et al., 2013). While our results are consistent with this notion of fronto-parietal involvement in dual n-back training effects, it is worth noting that our frontal alpha results are present in slightly more anterior clusters of electrodes than are typically seen. Given the nature of EEG, it is not possible to discern whether our results have distinct neural sources from previous work, but it is an interesting idea that may be relevant for future investigations working in source-space.

One interesting facet of our results is the lack of behavioral improvement on the Location and Relation WM tasks. The significant increase in frontal alpha power for the DNBT group was present despite no significant performance improvement on the tasks performed during the EEG recording. While these results may seem at odds with one another, the lack of behavioral change is also a benefit in interpreting the EEG changes. This suggests that our results are not simply driven by differences in behavior. At least one previous study has also found similar dissociations between changes in neural activity following training in the absence of changes in behavior (Vartanian et al., 2013). Somewhat surprisingly, we did not find any association between the changes in alpha power and the behavioral near transfer measures. In other words, there was no correlation between transfer to the object n-back task and changes in frontal alpha power during location or relation WM for the DNBT group. An important future direction will be to measure EEG activity while participants perform near transfer tasks to potentially pinpoint the mechanism underlying near transfer.

One additional limitation of the current study was the use of individual tasks as transfer measures. There has been a push in the field to examine transfer to construct level variables opposed to individual tests (Chooi & Thompson, 2012; Colom et al., 2013; Shipstead et al., 2012). It remains difficult to assess constructs with multiple measures and keep the testing sessions at a reasonable length. However, a critical future direction for the WM training field will be to systematically test for transfer to construct level variables and simultaneously uncover the potential neural basis of that transfer.

In summary, we sought to find what distinguishes n-back and complex span WM training tasks and how the two types of training compare with respect to near and far transfer. Our original hypothesis that relational WM was the critical factor differentiating these two forms of WM training was ultimately not supported. However, we found clear evidence that the effects of these two types of training are indeed distinct. In addition to improving behavioral performance on a different n-back task, DNBT was found to elicit significant changes in frontal alpha power during two other substantially different WM tasks. SST and our active control groups did not demonstrate any training-related changes in EEG measures. Thus, DNBT training appears to be a promising method for improving WM, but future work is

needed to test alternative hypotheses about what "active ingredients" of the n-back task make it a more effective training tool in order to further improve training protocols.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

Ackerman CM, Courtney SM. Spatial relations and spatial locations are dissociated within prefrontal and parietal cortex. Journal of Neurophysiology. 2012; 108(9):2419–2429. DOI: 10.1152/jn. 01024.2011 [PubMed: 22896722]

Au J, Sheehan E, Tsai N, Duncan GJ, Buschkuehl M, Jaeggi SM. Improving fluid intelligence with training on working memory: a meta-analysis. Psychon Bull Rev. 2015; 22(2):366–377. DOI: 10.3758/s13423-014-0699-x [PubMed: 25102926]

Badre D. Cognitive control, hierarchy, and the rostro-caudal organization of the frontal lobes. Trends in Cognitive Sciences. 2008; 12(5):193–200. DOI: 10.1016/J.Tics.2008.02.004 [PubMed: 18403252]

Bahlmann J, Blumenfeld RS, D'Esposito M. The Rostro-Caudal Axis of Frontal Cortex Is Sensitive to the Domain of Stimulus Information. Cereb Cortex. 2014; doi: 10.1093/cercor/bht419

Bastiaansen, M., Mazaheri, A., Jensen, O. Beyond ERP's: oscillatory neuronal dynamics. In: Luck, SJ., Kappenman, ES., editors. The Oxford Handbook of Event-Related Potential Components. USA: Oxford University Press; 2012.

Beatty EL, Jobidon ME, Bouak F, Nakashima A, Smith I, Lam Q, et al. Vartanian O. Transfer of training from one working memory task to another: behavioural and neural evidence. Front Syst Neurosci. 2015; 9:86.doi: 10.3389/fnsys.2015.00086 [PubMed: 26082694]

Berryhill ME, Peterson DJ, Jones KT, Stephens JA. Hits and misses: leveraging tDCS to advance cognitive research. Front Psychol. 2014; 5:800.doi: 10.3389/fpsyg.2014.00800 [PubMed: 25120513]

Blacker KJ, Courtney SM. Distinct neural substrates for maintaining spatial locations and relations in working memory. Frontiers in Human Neuroscience. 2016; 10:594.doi: 10.3389/fnhum.2016.00594 [PubMed: 27932963]

Blacker KJ, Ikkai A, Lakshmanan BM, Ewen JB, Courtney SM. The role of alpha oscillations in deriving and maintaining spatial relations in working memory. Cognitive Affective & Behavioral Neuroscience. 2016; 16(5):888–901. DOI: 10.3758/s13415-016-0439-y

Blacker KJ, Weisberg SM, Newcombe NS, Courtney SM. Keeping track of where we are: Spatial working memory in navigation. Visual Cognition. 2017; :1–12. DOI: 10.1080/13506285.2017.1322652

Bonnefond M, Jensen O. Alpha oscillations serve to protect working memory maintenance against anticipated distracters. Curr Biol. 2012; 22(20):1969–1974. DOI: 10.1016/j.cub.2012.08.029 [PubMed: 23041197]

Boot WR, Blakely DP, Simons DJ. Do action video games improve perception and cognition? Frontiers in Psychology. 2011; 2(226):1–6. DOI: 10.3389/fpsyg.2011.00226 [PubMed: 21713130]

Boot WR, Simons DJ, Stothart C, Stutts C. The pervasive problem of placebos in psychology: Why active control groups are not sufficient to rule out placebo effects. Perspectives on Psychological Science. 2013; 8(4):445–454. DOI: 10.1177/1745691613491271 [PubMed: 26173122]

Brainard DH. The Psychophysics Toolbox. Spat Vis. 1997; 10(4):433–436. [PubMed: 9176952]

Buschkuehl M, Hernandez-Garcia L, Jaeggi SM, Bernard JA, Jonides J. Neural effects of short-term training on working memory. Cogn Affect Behav Neurosci. 2014; 14(1):147–160. DOI: 10.3758/s13415-013-0244-9 [PubMed: 24496717]

Carpenter PA, Just MA, Shell P. What one intelligence test measures: a theoretical account of the processing in the Raven Progressive Matrices Test. Psychol Rev. 1990; 97(3):404–431. [PubMed: 2381998]

Chein JM, Morrison AB. Expanding the mind's workspace: Training and transfer effects with a complex working memory span task. Psychonomic Bulletin & Review. 2010; 17(2):193–199. DOI: 10.3758/PBR.17.2.193 [PubMed: 20382919]

Chiesa A, Calati R, Serretti A. Does mindfulness training improve cognitive abilities? A systematic review of neuropsychological findings. Clinical Psychology Review. 2011; 31(3):449–464. DOI: 10.1016/j.cpr.2010.11.003 [PubMed: 21183265]

Chooi WT, Thompson LA. Working memory training does not improve intelligence in healthy young adults. Intelligence. 2012; 40(6):531–542. DOI: 10.1016/J.Intell.2012.07.004

Cole MW, Bagic A, Kass R, Schneider W. Prefrontal Dynamics Underlying Rapid Instructed Task Learning Reverse with Practice. Journal of Neuroscience. 2010; 30(42):14245–14254. DOI: 10.1523/Jneurosci.1662-10.2010 [PubMed: 20962245]

Colom R, Roman FJ, Abad FJ, Shih PC, Privado J, Froufe M, et al. Jaeggi SM. Adaptive n-back training does not improve fluid intelligence at the construct level: Gains on individual tests suggest that training may enhance visuospatial processing. Intelligence. 2013; 41(5):712–727. DOI: 10.1016/J.Intell.2013.09.002

Conway AR, Kane MJ, Bunting MF, Hambrick DZ, Wilhelm O, Engle RW. Working memory span tasks: A methodological review and user's guide. Psychon Bull Rev. 2005; 12(5):769–786. [PubMed: 16523997]

Cusack R, Lehmann M, Veldsman M, Mitchell DJ. Encoding strategy and not visual working memory capacity correlates with intelligence. Psychonomic Bulletin & Review. 2009; 16(4):641–647. DOI: 10.3758/PBR.16.4.641 [PubMed: 19648446]

Engle RW, Tuholski SW, Laughlin JE, Conway ARA. Working memory, short-term memory, and general fluid intelligence. Journal of Experimental Psychology: General. 1999; 128(3):309–331. [PubMed: 10513398]

Fan J, McCandliss BruceD, Sommer T, Raz A, Posner MichaelI. Testing the efficiency and independence of attentional networks. Journal of Cognitive Neuroscience. 2002; 14(3):340–347. DOI: 10.1162/089892902317361886 [PubMed: 11970796]

Foster JJ, Sutterer DW, Serences JT, Vogel EK, Awh E. The topography of alpha-band activity tracks the content of spatial working memory. Journal of Neurophysiology. 2016; 115(1):168–177. DOI: 10.1152/jn.00860.2015 [PubMed: 26467522]

Green, CS., Gorman, T., Bavelier, D. Action Video-Game Training and Its Effects on Perception and Attentional Control. In: Strobach, T., Karbach, J., editors. Cognitive Training. Switzerland: Springer International Publishing; 2016. p. 107-116.

Haatveit BC, Sundet K, Hugdahl K, Ueland T, Melle I, Andreassen OA. The validity of d prime as a working memory index: results from the "Bergen n-back task". Journal of Clinical and Experimental Neuropsychology. 2010; 32(8):871–880. DOI: 10.1080/13803391003596421 [PubMed: 20383801]

Haegens S, Osipova D, Oostenveld R, Jensen O. Somatosensory working memory performance in humans depends on both engagement and disengagement of regions in a distributed network. Hum Brain Mapp. 2010; 31(1):26–35. DOI: 10.1002/hbm.20842 [PubMed: 19569072]

Herrmann CS, Senkowski D, Rottger S. Phase-locking and amplitude modulations of EEG alpha: Two measures reflect different cognitive processes in a working memory task. Exp Psychol. 2004; 51(4):311–318. DOI: 10.1027/1618-3169.51.4.311 [PubMed: 15620232]

Hillman CH, Erickson KI, Kramer AF. Be smart, exercise your heart: exercise effects on brain and cognition. Nature Reviews Neuroscience. 2008; 9(1):58–65. DOI: 10.1038/nrn2298 [PubMed: 18094706]

Holmes J, Gathercole SE, Dunning DL. Adaptive training leads to sustained enhancement of poor working memory in children. Developmental Science. 2009; 12(4):F9–F15. DOI: 10.1111/j. 1467-7687.2009.00848.x [PubMed: 19635074]

Hossiep, R., Turck, D., Hasella, M. Bochumer Matrizentest (BOMAT) advanced-short version. Gottingen: Hogrefe; 1999.

Ikkai A, Blacker KJ, Lakshmanan BM, Ewen JB, Courtney SM. Maintenance of relational information in working leads to suppression of the sensory cortex. Journal of Neurophysiology. 2014; 112(8): 1903–1915. DOI: 10.1152/jn.00134.2014 [PubMed: 25031260]

Jaeggi SM, Buschkuehl M, Jonides J, Perrig WJ. Improving fluid intelligence with training on working memory. Proceedings of the National Academy of Science. 2008; 105(19):6829–6833. doi: 10.1073pnas.0801268105.

Jaeggi SM, Buschkuehl M, Perrig WJ, Meier B. The concurrent validity of the N-back task as a working memory measure. Memory. 2010; 18(4):394–412. DOI: 10.1080/09658211003702171 [PubMed: 20408039]

Jaeggi SM, Buschkuehl M, Shah P, Jonides J. The role of individual differences in cognitive training and transfer. Memory & Cognition. 2014; 42(3):464–480. DOI: 10.3758/s13421-013-0364-z [PubMed: 24081919]

Jaeggi SM, Seewer R, Nirkko AC, Eckstein D, Schroth G, Groner R, Gutbrod K. Does excessive memory load attenuate activation in the prefrontal cortex? Load-dependent processing in single and dual tasks: functional magnetic resonance imaging study. NeuroImage. 2003; 19(2 Pt 1):210–225. [PubMed: 12814572]

Jaeggi SM, Studer-Luethi B, Buschkuehl M, Su YF, Jonides J, Perrig WJ. The relationship between n-back performance and matrix reasoning - implications for training and transfer. Intelligence. 2010; 38(6):625–635. DOI: 10.1016/J.Intell.2010.09.001

Jausovec N, Jausovec K. Working memory training: improving intelligence--changing brain activity. Brain Cogn. 2012; 79(2):96–106. DOI: 10.1016/j.bandc.2012.02.007 [PubMed: 22475577]

Jensen O, Gelfand J, Kounios J, Lisman JE. Oscillations in the alpha band (9-12 Hz) increase with memory load during retention in a short-term memory task. Cereb Cortex. 2002; 12(8):877–882. [PubMed: 12122036]

Jokisch D, Jensen O. Modulation of gamma and alpha activity during a working memory task engaging the dorsal or ventral stream. J Neurosci. 2007; 27(12):3244–3251. DOI: 10.1523/JNEUROSCI.5399-06.2007 [PubMed: 17376984]

Jonides J, Schumacher EH, Smith EE, Lauber EJ, Awh E, Minoshima S, Koeppe RA. Verbal working memory load affects regional brain activation as measured by PET. Journal of Cognitive Neuroscience. 1997; 9(4):462–475. DOI: 10.1162/Jocn.1997.9.4.462 [PubMed: 23968211]

Kaiser J, Heidegger T, Wibral M, Altmann CF, Lutzenberger W. Alpha synchronization during auditory spatial short-term memory. NeuroReport. 2007; 18(11):1129–1132. DOI: 10.1097/WNR. 0b013e32821c553b [PubMed: 17589312]

Kane MJ, Hambrick DZ, Tuholski SW, Wilhelm O, Payne TW, Engle RW. The Generality of Working Memory Capacity: A Latent-Variable Approach to Verbal and Visuospatial Memory Span and Reasoning. Journal of Experimental Psychology: General. 2004; 133(2):189–217. DOI: 10.1037/0096-3445.133.2.189 [PubMed: 15149250]

Kelly SP, Lalor EC, Reilly RB, Foxe JJ. Increases in alpha oscillatory power reflect an active retinotopic mechanism for distracter suppression during sustained visuospatial attention. Journal of Neurophysiology. 2006; 95(6):3844–3851. DOI: 10.1152/jn.01234.2005 [PubMed: 16571739]

Klimesch W, Sauseng P, Hanslmayr S. EEG alpha oscillations: the inhibition-timing hypothesis. Brain Res Rev. 2007; 53(1):63–88. DOI: 10.1016/j.brainresrev.2006.06.003 [PubMed: 16887192]

Klingberg T, Fernell E, Olesen PJ, Johnson M, Gustafsson P, Dahlstrom K, et al. Westerberg H. Computerized training of working memory in children with ADHD- A randomized, controlled trial. Journal of the American Academy of Child and Adolescent Psychiatry. 2005; 44(2):177–186. DOI: 10.1097/00004583-200502000-00010 [PubMed: 15689731]

Kundu B, Sutterer DW, Emrich SM, Postle BR. Strengthened effective connectivity underlies transfer of working memory training to tests of short-term memory and attention. J Neurosci. 2013; 33(20): 8705–8715. DOI: 10.1523/JNEUROSCI.5565-12.2013 [PubMed: 23678114]

Leung HC, Seelig D, Gore JC. The effect of memory load on cortical activity in the spatial working memory circuit. Cogn Affect Behav Neurosci. 2004; 4(4):553–563. [PubMed: 15849897]

Li CH, He X, Wang YJ, Hu Z, Guo CY. Visual Working Memory Capacity Can Be Increased by Training on Distractor Filtering Efficiency. Frontiers in psychology. 2017; 8

Li SC, Schmiedek F, Huxhold O, Rocke C, Smith J, Lindenberger U. Working memory plasticity in old age: practice gain, transfer, and maintenance. Psychol Aging. 2008; 23(4):731–742. DOI: 10.1037/a0014343 [PubMed: 19140644]

Libby LA, Hannula DE, Ranganath C. Medial Temporal Lobe Coding of Item and Spatial Information during Relational Binding in Working Memory. J Neurosci. 2014; 34(43):14233–14242. DOI: 10.1523/JNEUROSCI.0655-14.2014 [PubMed: 25339737]

Linke AC, Vincente-Grabovetsy A, Mitchell DJ, Cusack R. Encoding strategy accounts for individual differences in change detection measures of VSTM. Neuropsychologia. 2011; 49(6):1476–1486. DOI: 10.1016/j.neuropsychologia.2010.11.034 [PubMed: 21130789]

Lövdén M, Bäckman L, Lindenberger U, Schaefer S, Schmiedek F. A theoretical framework for the study of adult cognitive plasticity. Psychological Bulletin. 2010; 136(4):659.doi: 10.1037/a0020080 [PubMed: 20565172]

Luck SJ, Vogel EK. The capacity of visual working memory for features and conjunctions. Nature. 1997; 390(6657):279–281. DOI: 10.1038/36846 [PubMed: 9384378]

Lutz A, Slagter HA, Dunne JD, Davidson RJ. Attention regulation and monitoring in meditation. Trends Cogn Sci. 2008; 12(4):163–169. DOI: 10.1016/j.tics.2008.01.005 [PubMed: 18329323]

Maris E, Oostenveld R. Nonparametric statistical testing of EEG- and MEG-data. J Neurosci Methods. 2007; 164(1):177–190. DOI: 10.1016/j.jneumeth.2007.03.024 [PubMed: 17517438]

McKendrick R, Ayaz H, Olmstead R, Parasuraman R. Enhancing dual-task performance with verbal and spatial working memory training: continuous monitoring of cerebral hemodynamics with NIRS. NeuroImage. 2014; 85 Pt 3:1014–1026. DOI: 10.1016/j.neuroimage.2013.05.103 [PubMed: 23727530]

Miyake A, Friedman NP, Emerson MJ, Witzki AH, Howerter A, Wager TD. The unity and diversity of executive functions and their contributions to complex "Frontal Lobe" tasks: a latent variable analysis. Cognitive Psychology. 2000; 41(1):49–100. DOI: 10.1006/cogp.1999.0734 [PubMed: 10945922]

Morrison AB, Chein JM. Does working memory training work? The promise and challenges of enhancing cognition by training working memory. Psychonomic Bulletin and Review. 2011; 18:46–60. DOI: 10.3758/s13423-010-0034 [PubMed: 21327348]

Nichols TE, Holmes AP. Nonparametric permutation tests for functional neuroimaging: a primer with examples. Hum Brain Mapp. 2002; 15(1):1–25. [PubMed: 11747097]

Parkin BL, Ekhtiari H, Walsh VF. Non-invasive Human Brain Stimulation in Cognitive Neuroscience: A Primer. Neuron. 2015; 87(5):932–945. DOI: 10.1016/j.neuron.2015.07.032 [PubMed: 26335641]

Pelli DG. The VideoToolbox software for visual psychophysics: transforming numbers into movies. Spat Vis. 1997; 10(4):437–442. [PubMed: 9176953]

Percival, D., Walden, A. Spectral analysis for physical applications. Cambridge: Cambridge University Press; 1993.

Powers KL, Brooks PJ, Aldrich NJ, Palladino MA, Alfieri L. Effects of video-game play on information processing: A meta-analytic investigation. Psychonomic Bulletin & Review. 2013; 20(6):1055–1079. DOI: 10.3758/s13423-013-0418-z [PubMed: 23519430]

Redick TS, Broadway JM, Meier ME, Kuriakose PS, Unsworth Nash, Kane MJ, Engle RandallW. Measuring working memory capacity with automated complex span tasks. European Journal of Psychological Assessment. 2012; 28(3):164–171. DOI: 10.1027/1015-5759/a000123

Redick TS, Lindsey DR. Complex span and n-back measures of working memory: a meta-analysis. Psychon Bull Rev. 2013; 20(6):1102–1113. DOI: 10.3758/s13423-013-0453-9 [PubMed: 23733330]

Redick TS, Shipstead Z, Harrison TL, Hicks KL, Fried DE, Hambrick DZ, et al. Engle RW. No evidence of intelligence improvement after working memory training: a randomized, placebo-

controlled study. Journal of Experimental Psychology-General. 2013; 142(2):359–379. DOI: 10.1037/a0029082 [PubMed: 22708717]

Richmond LL, Morrison AB, Chein JM, Olson IR. Working memory training and transfer in older adults. Psychol Aging. 2011; 26(4):813–822. DOI: 10.1037/a0023631 [PubMed: 21707176]

Rogers RD, Monsell S. Costs of a Predictable Switch between Simple Cognitive Tasks. Journal of Experimental Psychology-General. 1995; 124(2):207–231. DOI: 10.1037//0096-3445.124.2.207

Roux F, Uhlhaas PJ. Working memory and neural oscillations: alpha-gamma versus theta-gamma codes for distinct WM information? Trends Cogn Sci. 2013; doi: 10.1016/j.tics.2013.10.010

Rudebeck SR, Bor D, Ormond A, O'Reilly JX, Lee ACH. A Potential Spatial Working Memory Training Task to Improve Both Episodic Memory and Fluid Intelligence. PLoS One. 2012; 7(11)doi: 10.1371/journal.pone.0050431

Sadaghiani S, Scheeringa R, Lehongre K, Morillon B, Giraud AL, D'Esposito M, Kleinschmidt A. alpha-band phase synchrony is related to activity in the fronto-parietal adaptive control network. J Neurosci. 2012; 32(41):14305–14310. DOI: 10.1523/JNEUROSCI.1358-12.2012 [PubMed: 23055501]

Salminen T, Kuhn S, Frensch PA, Schubert T. Transfer after Dual n-Back Training Depends on Striatal Activation Change. J Neurosci. 2016; 36(39):10198–10213. DOI: 10.1523/JNEUROSCI. 2305-15.2016 [PubMed: 27683914]

Sauseng P, Klimesch W, Schabus M, Doppelmayr M. Fronto-parietal EEG coherence in theta and upper alpha reflect central executive functions of working memory. International Journal of Psychophysiology. 2005; 57(2):97–103. DOI: 10.1016/j.ijpsycho.2005.03.018 [PubMed: 15967528]

Sauseng P, Klimesch W, Stadler W, Schabus M, Doppelmayr M, Hanslmayr S, et al. Birbaumer N. A shift of visual spatial attention is selectively associated with human EEG alpha activity. Eur J Neurosci. 2005; 22(11):2917–2926. DOI: 10.1111/j.1460-9568.2005.04482.x [PubMed: 16324126]

Schneiders JA, Opitz B, Krick CM, Mecklinger A. Separating intra-modal and across-modal training effects in visual working memory: an fMRI investigation. Cereb Cortex. 2011; 21(11):2555–2564. DOI: 10.1093/cercor/bhr037 [PubMed: 21471559]

Shipstead Z, Redick TS, Engle RW. Is Working Memory Training Effective? Psychological Bulletin. 2012; 138(4):628–654. DOI: 10.1037/A0027473 [PubMed: 22409508]

Simons DJ, Boot WR, Charness N, Gathercole SE, Chabris CF, Hambrick DZ, Stine-Morrow EA. Do "Brain-Training" Programs Work? Psychol Sci Public Interest. 2016; 17(3):103–186. DOI: 10.1177/1529100616661983 [PubMed: 27697851]

Stephenson CL, Halpern DF. Improved matrix reasoning is limited to training on tasks with a visuospatial component. Intelligence. 2013; 41:341–357. DOI: 10.1016/j.intell.2013.05.006

Tang YY, Holzel BK, Posner MI. The neuroscience of mindfulness meditation. Nature Reviews Neuroscience. 2015; 16(4):213–225. DOI: 10.1038/nrn3916 [PubMed: 25783612]

Thompson TW, Waskom ML, Gabrieli JD. Intensive Working Memory Training Produces Functional Changes in Large-scale Frontoparietal Networks. J Cogn Neurosci. 2016; :1–14. DOI: 10.1162/jocn_a_00916

Thompson TW, Waskom ML, Garel KL, Cardenas-Iniguez C, Reynolds GO, Winter R, et al. Gabrieli JD. Failure of working memory training to enhance cognition or intelligence. PLoS One. 2013; 8(5):e63614.doi: 10.1371/journal.pone.0063614 [PubMed: 23717453]

Todd, JJay, Marois, Rene. Capacity limit of visual short-term memory in human posterior parietal cortex. Nature. 2004; 428(6984):751–754. DOI: 10.1038/nature02466 [PubMed: 15085133]

Vanderplas JM, Garvin EA. The association value of random shapes. J Exp Psychol. 1959; 57(3):147–154. [PubMed: 13641586]

Vartanian O, Jobidon ME, Bouak F, Nakashima A, Smith I, Lam Q, Cheung B. Working memory training is associated with lower prefrontal cortex activation in a divergent thinking task. Neuroscience. 2013; 236:186–194. DOI: 10.1016/j.neuroscience.2012.12.060 [PubMed: 23357116]
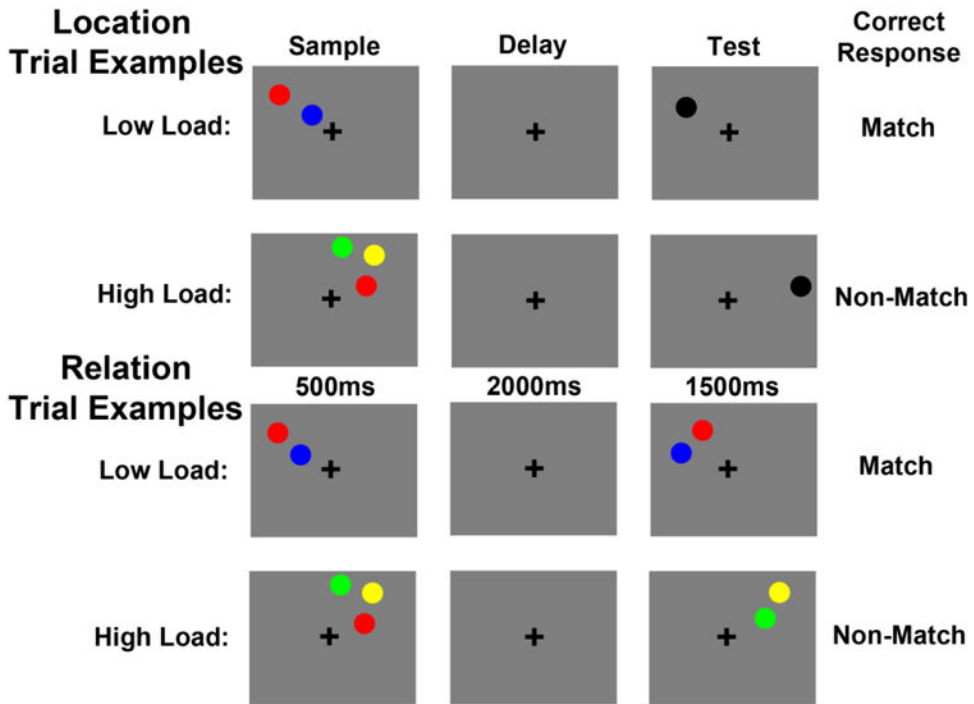
Zanto TP, Rubens MT, Thangavel A, Gazzaley A. Causal role of the prefrontal cortex in top-down modulation of visual processing and working memory. Nat Neurosci. 2011; 14(5):656–661. DOI: 10.1038/nn.2773 [PubMed: 21441920]
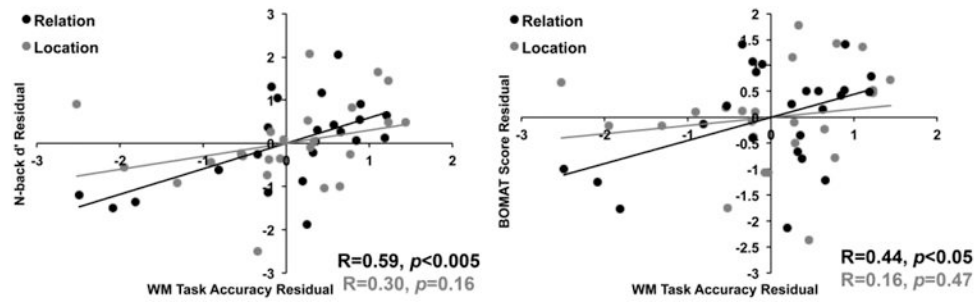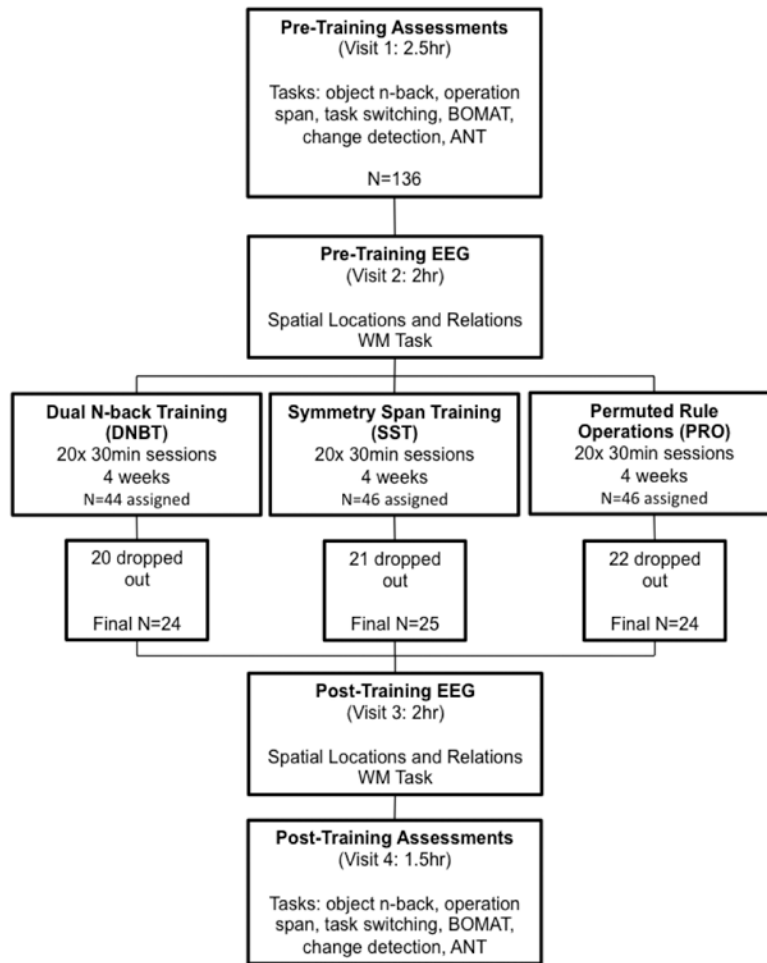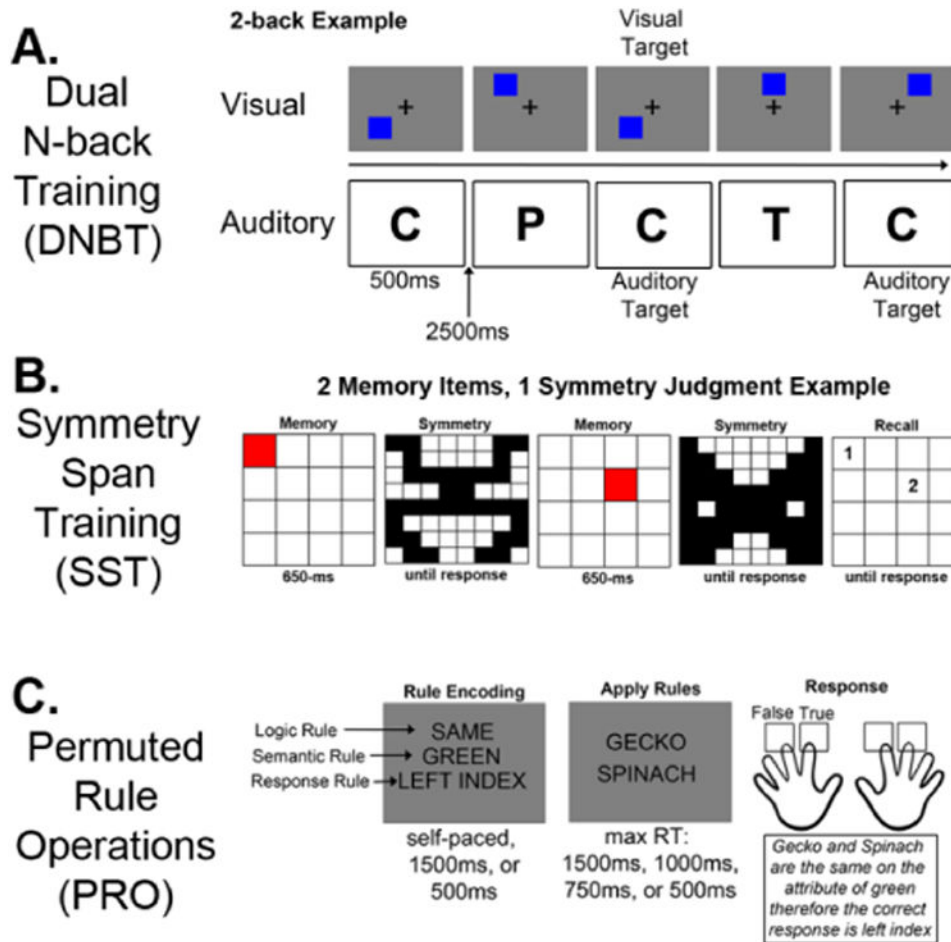
Figure 1.
Trial examples for the Spatial Locations and Relations task. Under low load, Location trials required participants to imagine a line between two sample circles, hold the location of that line in memory across a delay and then decide if a test circle fell in that location or not. Under high load, Location trials required participants to maintain the locations of three circles in memory and then decide if a test circle fell in one of those locations or in a completely new location. Under low load, Relation trials required participants to maintain the vertical relationship (above/below) of two sample circles and then decide if two test circles were in the same relationship. Under high load, Relation trials required participants to maintain the three vertical relationships between three sample circles and then decide if one of those pairs were presented in the same relationship at test.

Figure 2.
Partial correlations between Location and Relation WM accuracy and the n-back and BOMAT. Relation WM accuracy significantly predicted n-back and BOMAT performance, while controlling for Location WM accuracy.

Figure 3.
General study procedures including sample size and attrition rate by group.

Figure 4.
Task schematics for each of the three training tasks: A) Dual n-back training, B) Symmetry span training, and C) Permuted rule operations.
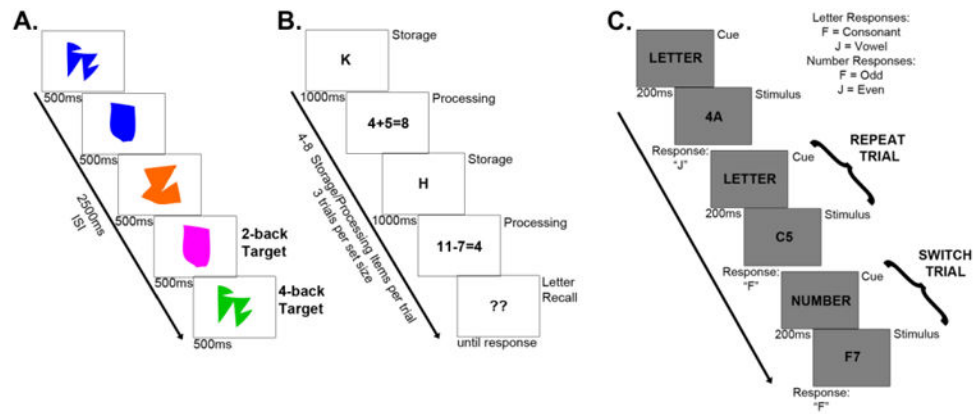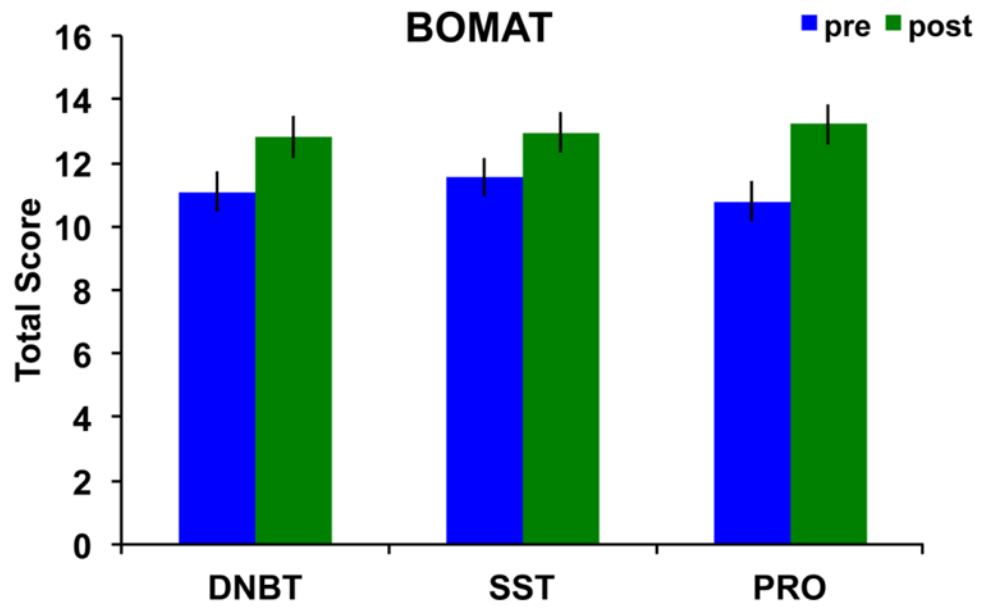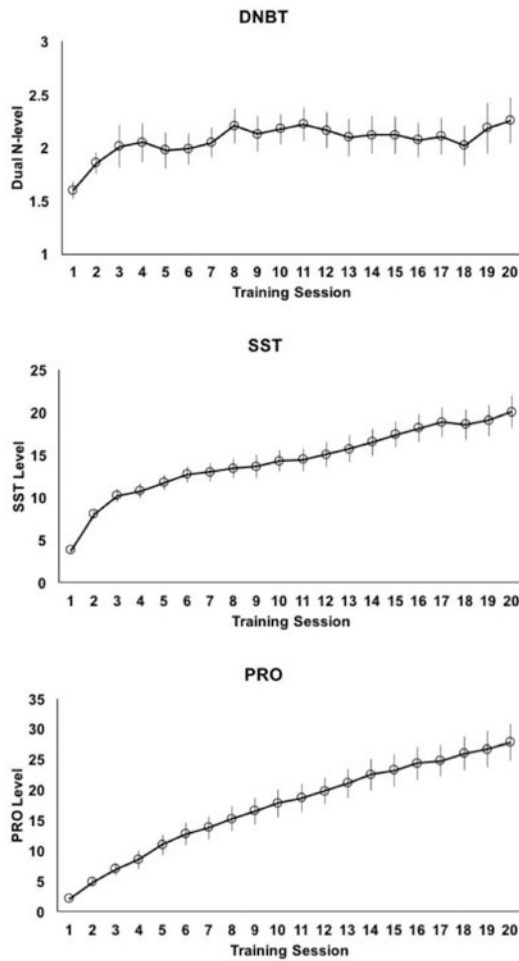
Figure 5.
Near transfer assessment tasks. A) Trial schematic for the object n-back task illustrating example 2- and 4-back target scenarios. Participants completed separate blocks of each n-level (1-5). B) Trial schematic for the Ospan task. Participants completed set sizes 4-8. C) Trial schematics for the Task Switching task illustrating example "repeat" and "switch" trials.
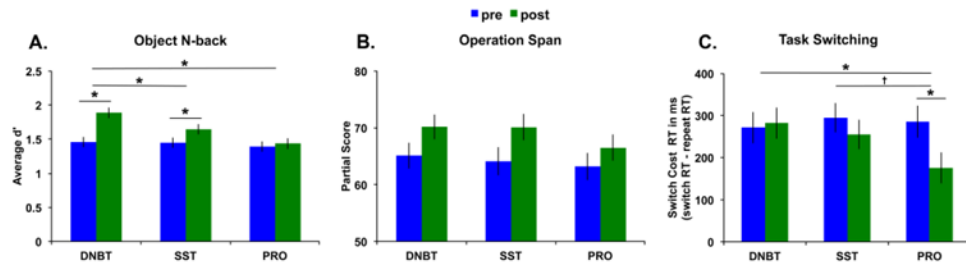
Figure 6.

Near transfer results for each training group. A) Significant near transfer to the object n-back task for the DNBT group. B) The data for the OSpan task were in the expected direction with numerically greater gains for the SST group, but the results did not reach significance. C) Significant near transfer to the task switching paradigm for the PRO group. Error bars represent standard error of the mean. *$p<0.05$, †$p=0.1$.
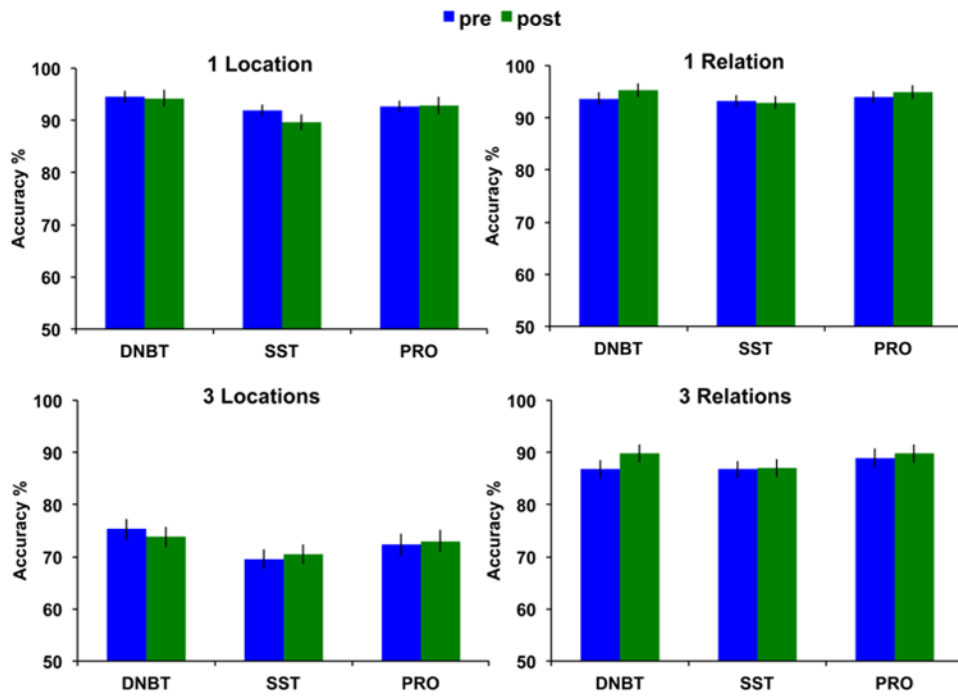
Figure 7.
BOMAT performance before and after training by group showing no evidence of significant transfer for any specific group. Error bars represent standard error of the mean.
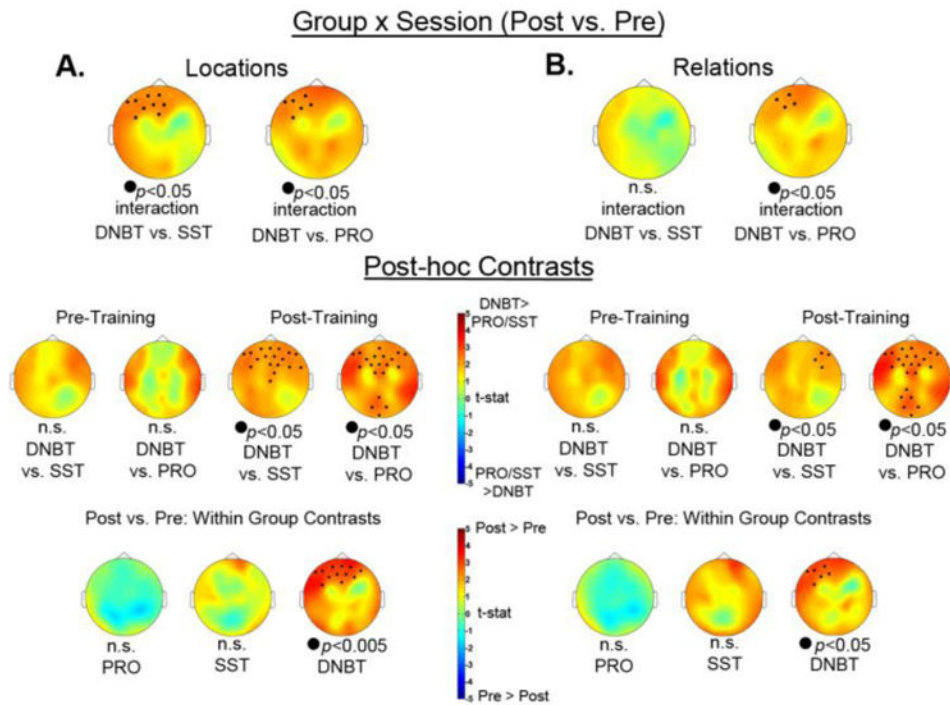
Figure 8.
Group average training performance for each group. Data is shown as the average level achieved per session. For the DNBT group, data is shown by n-level. For the SST group, every third level added an additional memory item, whereas in between levels added symmetry judgments in between memory items. For the PRO group, each level required faster encoding of the rules and then every fourth level added in a new rule, then starting with level 31 the amount of time to respond decreased. Error bars represent standard error of the mean.
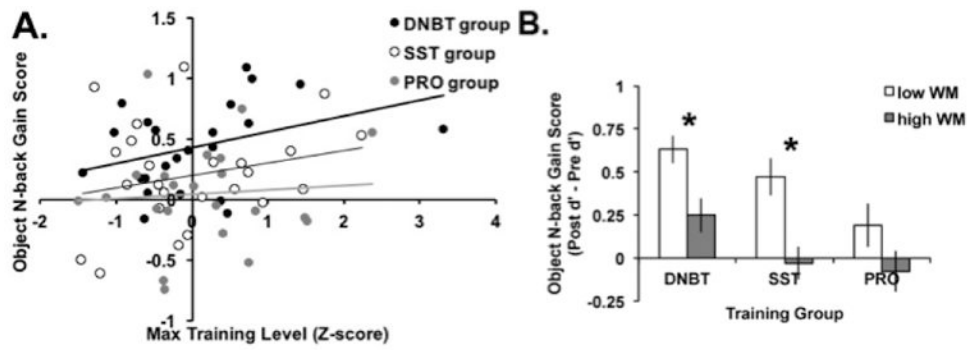
Figure 9.
Predictors of near transfer to the object n-back task. A) Scatterplot for each training group illustrating the relationship between standardized maximum training level reached and improvement on the object n-back task. B) Baseline performance on the object n-back task was associated with greater gains after training on that same task for the WM groups but not the PRO group. Error bars represent standard error of the mean. *$p<0.05$.

Figure 10.
Behavioral accuracy data for the Relation and Location WM task that participants completed while EEG was recorded. Error bars represent standard error of the mean.

Figure 11.
Nonparametric permutation test results for delay period alpha power (8-13Hz). Results are shown for group × session interactions and post-hoc contrasts separately for Location (A) and Relation (B) trials.

Table 1

Descriptive statistics for accuracy in the Locations/Relations WM task in Experiment 1.

Descriptive Statistics	Location Trials	Relation Trials
Low Load	*M*=93.8% (*SD*=6.1)	*M*=93.6% (*SD*=4.7)
High Load	*M*=82.9% (*SD*=9.7)	*M*=86.5% (*SD*=11.2)

Table 2

Test-Retest reliability results for the transfer measures of interest.

Test-Retest Reliability for Near Transfer Measures	
Measure	**Cronbach's alpha**
Object n-back: Average d′	0.65
Ospan: Partial Score	0.88
Task Switching: Switch Cost	0.72
BOMAT	0.72