

Running head: WORKING MEMORY, DIVERGENT THINKING, AND NEURAL EFFICIENCY

Working Memory Training Is Associated with Lower Prefrontal Cortex Activation in a Divergent
Thinking Task

Oshin Vartanian^{1,2}, Marie-Eve Jobidon¹, Fethi Bouak¹, Ann Nakashima¹, Ingrid Smith¹, Quan
Lam¹, and Bob Cheung¹

1 Defence R&D Canada—Toronto

2 University of Toronto—Scarborough

Address for correspondence:

Oshin Vartanian, Ph.D.
Defence R&D Canada—Toronto
1133 Sheppard Avenue West
P.O. Box 2000
Toronto, ON
M3M 3B9 Canada
Email: oshin.vartanian@drdc-rddc.gc.ca

Abstract

Working memory (WM) training has been shown to lead to improvements in WM capacity and fluid intelligence. Given that divergent thinking loads on WM and fluid intelligence, we tested the hypothesis that WM training would improve performance and moderate neural function in the Alternate Uses Task (AUT)—a classic test of divergent thinking. We tested this hypothesis by administering the AUT in the functional magnetic resonance imaging scanner following a short regimen of WM training (experimental condition), or engagement in a choice reaction time task not expected to engage WM (active control condition). Participants in the experimental group exhibited significant improvement in performance in the WM task as a function of training, as well as a significant gain in fluid intelligence. Although the two groups did not differ in their performance on the AUT, activation was significantly lower in the experimental group in ventrolateral prefrontal and dorsolateral prefrontal cortex—two brain regions known to play dissociable and critical roles in divergent thinking. Furthermore, gain in fluid intelligence mediated the effect of training on brain activation in ventrolateral prefrontal cortex. These results indicate that a short regimen of WM training is associated with lower prefrontal activation—a marker of neural efficiency—in divergent thinking.

Keywords: Working memory, training, creativity, intelligence.

1 Introduction

Divergent thinking is defined as thinking that can lead in various directions (Guilford, 1967; Runco, 1999). Historically, divergent thinking ability has been measured using tasks that instruct participants to generate multiple solutions to open-ended problems. For example, a classic task from the divergent thinking literature is the Alternate Uses Task (AUT), in which participants are asked to generate as many uses as possible in response to prompts of common objects (e.g., brick). These responses can then be scored in relation to various indices of interest, including fluency (number of responses), originality (statistical infrequency of responses), or flexibility (categorical variability of responses), among others. Divergent thinking tasks are among the most commonly used in studies of creativity because their scores are considered to reflect estimates of potential creative thought (Plucker and Renzulli, 1999), and are predictive of real-life creativity (Plucker, 1999). However, in accordance with Guilford's (1967) original formulation, divergent production should not be viewed as a singular index of creativity, but rather as one of the mental operations that contribute to the structure of intellect.

Recently, creativity researchers have begun to reexamine the link between divergent thinking on the one hand, and working memory (WM) and fluid intelligence on the other hand. It seems reasonable to suspect that divergent thinking would load on WM capacity, given that divergent thinking requires the maintenance and manipulation of relevant concepts in the span

of attention (Vartanian, 2011).¹ One would also expect divergent thinking to load on fluid intelligence—defined as the ability to adapt to new situations and to perceive new patterns and relationships (Cattell, 1963; Unsworth and Engle, 2005). Indeed, recent evidence has shown that divergent thinking requires executive functions in the service of interference resolution and strategy application (Gilhooly et al., 2007). In addition, there is now evidence to support the loading of divergent thinking on fluid intelligence (Nussbaum and Silvia, 2011), as well as the loading of creativity on fluid intelligence (Sligh et al., 2005). The data from this literature suggest that fluid intelligence (and executive functions) contribute to divergent thinking ability (and creativity).

Given the aforementioned findings relating divergent thinking to executive functions and fluid intelligence, our aim in this experiment was to test the hypothesis that a short but temporally concentrated WM training regimen would translate into better performance on a divergent thinking task. This hypothesis was motivated by two strands of evidence. First, a substantial body of evidence has accumulated to demonstrate that targeted and repeated WM training can increase WM capacity (for reviews see Klingberg, 2010; Morrison and Chein, 2011). To the extent that divergent thinking ability draws on WM capacity, increases in the latter should contribute to improvements in the former. Second, WM training has also been shown to improve fluid intelligence (Jaeggi et al., 2008, 2010; for a review see Buschkuhl and Jaeggi, 2010). This suggests that the benefits of WM training on divergent thinking could take two routes: either directly by increasing WM capacity, or indirectly by way of improving fluid

¹ WM capacity is necessary to establish attentional control for focused cognition, although creativity thrives as a function of flexible switching between focused and defocused modes of cognition in relation to task demands (Vartanian, 2009; Wiley and Jarosz, 2012).

intelligence. In the latter sense, WM training can be viewed as an intervention strategy to boost fluid intelligence in the service of improving divergent thinking.

In addition to the hypothesized behavioural effects, we were also interested in examining the neural system that would underlie the transfer effects. If successful transfer from WM training to divergent thinking is a function of the goodness-of-fit between capacities enhanced during WM training and the cognitive requirements of the target activity, then the neural localization of the transfer may be predicted based on the extent of overlap between brain systems known to underlie WM (training task), fluid intelligence, and divergent thinking (target of transfer task). There is now strong evidence to suggest that the ventral and dorsal aspects of the prefrontal cortex (PFC) underlie the expression of intelligence (Deary et al., 2010; Jung and Haier, 2007), WM (Baddeley, 2003; Prabhakaran et al., 2000), and divergent thinking (Fink et al., 2009; Goel and Vartanian, 2005). Thus, one would predict that the impact of WM training on divergent thinking (possibly mediated by fluid intelligence) should be localizable in the PFC.

However, neurally, the direction of the effect underlying transfer from WM training to target tasks remains unknown. This is because there are at least two different ways in which this effect could be realized in the PFC. On the one hand, there is reason to believe that improving fluid intelligence as a function of WM training should result in enhanced neural efficiency—operationalized here and elsewhere as a lower blood oxygenation-level dependent (BOLD) signal—in the PFC. This prediction is consistent with the inverse relation observed between fluid intelligence and metabolic rate in the parieto-frontal network using a variety of cognitive tasks that tap fluid intelligence (Deary et al., 2010; Jung and Haier, 2007; Neubauer et

al., 2002, 2005; see also van der Heuvel et al., 2009), as well as indications that this inverse relation is most likely to be observed in the frontal cortex (Neubauer and Fink, 2009). On the other hand, there is also evidence for a positive correlation between neural activity and fluid intelligence (Gray et al., 2003; Lee et al., 2006; Luders et al., 2009), as well as data demonstrating an increase in PFC activity following WM training (Olesen, 2004). In conjunction, the current evidence suggests that while the impact of WM training on fluid intelligence is likely to manifest itself in the PFC, the direction of this effect is difficult to predict.

Despite the heterogeneous nature of the findings linking WM training to fluid intelligence, Klingberg (2010) noted a consistent pattern such that short periods of WM training (<3 hours) had been shown to result in decreased brain activity, whereas long periods of WM training had been shown to result in an mixture of increased and decreased brain activity. He suggested that the decreases in activation can occur as a result of strategy learning, priming during encoding, and time-on-task effects. In turn, during longer WM training these reductions would be co-occurring with increases in capacity correlated with activity in the intraparietal cortex, middle and superior frontal gyri, and the caudate nucleus. Thus, duration of training may be an important moderator of the relation between WM training and brain activation. However, in their own review of the literature on the neural effects of WM training, Buschkuhl et al. (2012) argued against Klingberg's (2010) categorization of the effects based on training duration, calling for additional data to understand the impact of WM training on neural function.

1.1 Present study and hypotheses

In the present study we tested the effect of WM training on behavioural performance and neural function in the context of the AUT in the functional magnetic resonance imaging (fMRI) scanner. We also examined the possible mediating role of fluid intelligence in the transfer process. Practice durations and frequencies in previous studies involving WM training have varied greatly, ranging from one 20- or 30-minute session to 20 hours spread over 10 weeks (see Buschkuhl et al., 2012, Table 1; Klingberg, 2010, Table 2; see also Morrison and Chein, 2011). In contrast, studies involving WM training to improve fluid intelligence have typically employed short training sessions (i.e., \approx 17-25 minutes) administered with high frequency (8-20 sessions) (Jaeggi et al., 2008, 2010). We opted to test the impact of three closely spaced 20-minute training sessions on performance and neural function. Our decision to focus on a short and concentrated regimen of WM training was motivated by our desire to assess the feasibility of WM training as an intervention strategy in applied professional and educational settings where the implementation of lengthy training regimens might be impractical. We implemented a short time lag between the completion of WM training and the assessment of divergent thinking ability to maximize the likelihood of observing short-term transfer effects.

We tested four hypotheses: (1) we predicted an increase in WM capacity in participants enrolled in the experimental condition who trained on the n-back task; (2) compared to participants enrolled in an active control condition, we predicted that participants enrolled in the experimental condition would exhibit improvement in fluid intelligence; (3) compared to participants enrolled in an active control condition, we predicted better performance on the AUT for participants in the experimental condition; and, finally, (4) we expected to observe

variation in the BOLD response in the ventral and/or dorsal PFC as a function of WM training.

We were particularly interested in these two regions because ventral and dorsal right PFC had been linked to (a) generating responses and (b) frequency (i.e., number) of generated responses respectively in divergent thinking (Goel and Vartanian, 2005). Although the location of the transfer effect was specified a priori, the direction of that effect was unspecified given the variability observed in previous studies of WM training (Buschkuhl et al., 2012; Klingberg, 2010).

2 Experimental procedures

2.1 Participants

The participants were 34 neurologically healthy right-handed volunteers ($M = 30.79$ years, $SD = 7.06$) with normal or corrected-to-normal vision recruited from the Greater Toronto Area. Handedness was assessed using a standard self-report questionnaire (Oldfield, 1971). The participants were assigned randomly to the experimental ($N = 17$) and active control groups ($N = 17$), and matched for sex (13 females, 4 males) and age, $t(32) = .46$, $p = .65$, $d = .17$.²

2.2 Materials and procedure

2.2.1 Cognitive training

All participants completed three training sessions individually on separate days. Average lag time between successive sessions was 1.51 days ($SD = 3.75$). There was no significant difference between the two groups in lag time between sessions, $t(32) = 1.12$, $p = .27$, $d = .38$.

The duration of each session was 20 minutes (excluding time required to read the instructions,

² Throughout the manuscript Cohen's d is used as a measure of effect size for t tests, adjusted for whether the t test is based on dependent or independent samples (Cohen, 1988). We used the following online calculator for calculating d : <http://www.cognitiveflexibility.org/effectsize/>

which was self paced). Training was administered using the *Cognitive Test Software* (Grushcow, 2008) installed on laptops.

Participants in the experimental group completed the n-back task. Although both the n-back task and the dual n-back task have been used in previous training studies, they have been shown to have comparable efficacy in improving WM capacity and fluid intelligence (Jaeggi et al., 2010). Each session consisted of four blocks—two blocks of 2-back and two blocks of 3-back—administered in alternating order and always starting with 2-back. Each letter was presented for 500 ms. Inter-stimulus interval (ISI) was a blank screen, presented for 2500 ms. Participants pressed the spacebar whenever they detected a match.

Participants in the active control group completed the 4-choice reaction time (RT) task. On each trial of this task one of four adjacent locations on the computer screen was highlighted randomly. Participants were instructed to press one of four keys (1-4) corresponding to the highlighted location on each trial. We selected this task specifically for three reasons. First, it is not considered a WM task, but rather an attention task. Second, whereas performance on the n-back task should be predictive of fluid intelligence, performance on the 4-choice RT task should be unrelated to fluid intelligence. Third, based on previous data collected in our lab, we expected performance to be at ceiling in all sessions. Specifically, a previous study ($N = 20$) had demonstrated that accuracy across three sessions (one week apart) was 99% ($SD = 1$), 98% ($SD = 2$), and 97% ($SD = 3$) respectively (Nakashima et al., 2011). This would ensure that unlike in the experimental group that trained on the n-back task, cognitive exertion would be minimal for participants in the active control group completing the control task. Note that because of

the aforementioned features, this active control task should be construed as a condition controlling for task engagement.

2.2.2 Fluid intelligence

Our measure of fluid intelligence was *Raven's Advanced Progressive Matrices (RAPM)* (Raven et al., 1998). RAPM was divided into two parallel forms based on even and odd items (see Jaeggi et al., 2008). Each form contained 18 problems. Upon recruitment, one part was administered to calculate baseline fluid intelligence. The second part was administered following the completion of training to calculate gain. All participants were tested individually on the RAPM, and given 10 minutes to complete as many problems as possible.

2.2.3 Target tasks

Following training, participants completed the AUT in the fMRI scanner.³ There was no significant difference in lag time (days) between the completion of training and data collection in the fMRI scanner between the experimental ($M = 2.94$, $SD = 1.97$) and active control ($M = 3.18$, $SD = 1.29$) groups, $t(32) = .55$, $p = .59$, $d = .15$. The AUT is a classic divergent thinking task from the creativity literature (Guilford, 1967). The scanner version of the AUT was modeled after Fink et al. (2009). The task was presented in two blocks (i.e., uses and characteristics), the order of which was counterbalanced across participants. Each of the 20 trials in the *uses* block had the same structure. During the *generation* phase participants were presented with the name of a common object (e.g., knife) and instructed to think of as many uses for it as possible for 12000 ms. In this phase the name of the object appeared in black ink. The *response* phase

³ Two additional tasks (psychomotor vigilance task and delayed matching-to-sample task) were completed in the same scanning session in counterbalanced order. The results involving those tasks will be reported elsewhere.

followed immediately afterwards during which participants were given 3000 ms to enter the *number* of generated uses.⁴ In this phase the name of the object appeared in green. This color change acted as a prompt to enter the response as quickly as possible. This was followed by an inter-trial interval (ITI) varying randomly between 4,000-6,000 ms. Each trial in the *characteristics* block had an identical structure, except that participants were instructed not to generate uses, but instead to recall from long-term memory physical features characteristic of the object. For example, possible physical features for “knife” could be solid, sharp, metallic, etc.⁵

2.2.4 fMRI acquisition

A 3-Tesla MR scanner with an 8-channel head coil (Discovery MR750, 22.0 software, GE Healthcare, Waukesha, WI) was used to acquire T1 anatomical volume images (.86 × .86 × 1.0 mm voxels). For functional imaging, T2*-weighted gradient echo spiral-in/out acquisitions were used to produce 26 contiguous 5 mm thick axial slices (repetition time [TR] = 2000 ms; echo time [TE] = 30 ms; flip angle [FA] = 70°; field of view [FOV] = 200 mm; 64 × 64 matrix; voxel dimensions = 3.1 × 3.1 × 5.0 mm), positioned to cover the whole brain. The first five volumes were discarded to allow for T1 equilibration effects. The number of volumes acquired was 418.

2.2.5 fMRI analysis

⁴ Although responses to divergent thinking tasks can be scored in multiple ways, fluency (i.e., the number of generated solutions) accounts for the majority of variance in divergent thinking tasks (Plucker and Renzulli, 1999).

⁵ The characteristics block was included to mimic Fink et al.'s (2009) design as closely as possible. We had no hypotheses involving the effect of cognitive training on recall from long-term memory.

Data were analyzed using Statistical Parametric Mapping (SPM8) (www.fil.ion.ucl.ac.uk/spm/). Head movement was less than 2 mm in all cases. All functional volumes were spatially realigned to the first volume. A mean image created from realigned volumes was spatially normalized to the MNI EPI brain template using nonlinear basis functions. The derived spatial transformation was applied to the realigned T2* volumes, and spatially smoothed with an 8 mm full-width at half-maximum (FWHM) isotropic Gaussian kernel. Time series across each voxel were high-pass filtered with a cut-off of 128 s, using cosine functions to remove section-specific low frequency drifts in the BOLD signal. Condition effects at each voxel were estimated according to the general linear model (GLM) and regionally specific effects compared using linear contrasts. The BOLD signal was modeled as a box-car, convolved with a canonical hemodynamic response function. Each contrast produced a statistical parametric map consisting of voxels where the z -statistic was significant at $p < .001$. Using a random-effects analysis, reported activations survived voxel-level intensity threshold of $p < .05$, corrected for multiple comparisons using the Bonferroni family-wise (whole-brain) correction.

3 Results

3.1 Cognitive training

We conducted two separate within-subjects ANOVAs for the experimental and active control groups, with session (1, 2, 3) as the within-subjects variable and percentage accuracy on the trained task as the dependent variable. There was a significant effect of session for the experimental group, $F(2, 32) = 35.61$, $p < .000001$, partial $\eta^2 = .69$ (Figure 1). Exploring this effect further, paired t tests demonstrated that there was a significant increase in accuracy between session 1 and session 2 ($t[16] = 5.35$, $p = .000065$, $d = 1.30$) and between session 2 and

session 3 ($t[16] = 2.26, p = .038, d = .55$). In contrast, as predicted (see Nakashima et al., 2011), for the active control group performance was at ceiling across three sessions, $F(2, 32) = 2.34, p = .112, \text{partial } \eta^2 = .13$ (Figure 1).

3.2 Fluid intelligence

There was no significant difference between the experimental ($M = 10.35, SD = 2.62$) and active control groups ($M = 11.29, SD = 2.64$) in baseline RAPM scores, $t(32) = 1.04, p = .31, d = .36$. A mixed-model ANOVA demonstrated a significant group (experimental vs. active control) \times time point (pre- vs. post-training) interaction, $F(1, 32) = 5.90, p = .021, \text{partial } \eta^2 = .16$. Specifically, for the experimental group there was a significant improvement in RAPM scores following training ($M = 11.18, SD = 2.53$) compared to baseline, $t(16) = 2.46, p = .026, d = .60$. In contrast, in the active control group no change was observed following training ($M = 10.41, SD = 2.24$) compared to baseline, $t(16) = 1.43, p = .17, d = .35$.⁶

To test our theoretical prediction that performance on the n-back task should be related to fluid intelligence whereas performance on the 4-choice RT task should be unrelated to fluid intelligence, we ran two regressions in which we regressed baseline RAPM scores onto n-back or 4-choice RT scores (collected at session 1), separately depending on the condition in which participants were (experimental or active control). The results demonstrated that n-back performance linearly predicted RAPM scores, $r = .69, p = .002$, accounting for almost 47% of

⁶ Another way to conduct this analysis involves conducting an ANCOVA in which group (experimental vs. active control) is entered as the independent variable, RAPM scores following training as the dependent variable, and RAPM scores at baseline as the covariate. Consistent with the analysis reported above, this analysis also demonstrated a significant effect for group, $F(1, 31) = 4.54, p = .041, \text{partial } \eta^2 = .13$.

the observed variance in baseline fluid intelligence (linear R^2). In contrast 4-choice RT performance did not predict RAPM scores, $\beta = .17, p = .522$, accounting for less than 3% of the observed variance in baseline fluid intelligence (Figure 2). This adds weight to our task analysis and the argument that training on the n-back task is more likely to affect RAPM scores than engagement in the 4-choice task.

3.3 Target task

3.3.1 Behavioural effects

Average RT for generating uses was 1105 ms ($SD = 716$), and average number of generated uses was 4.61 ($SD = 1.76$). There was no significant difference in number of uses ($t[31] = .77, p = .45, d = .27$) or RT ($t[31] = .02, p = .98, d = .01$) between the experimental and active control groups. Average RT for recalling characteristics was 1364 ms ($SD = 629$), and the average number of recalled characteristics was 4.70 ($SD = 1.57$). There was no significant difference in number of recalled characteristics ($t[31] = .40, p = .69, d = .14$) or RT ($t[31] = .24, p = .81, d = .08$) between the two groups.

3.3.2 Neural effects

Using an event-related design, we specified regressors corresponding to the generation phase (coupled with the parameter indicating number of uses or characteristics generated—depending on the condition), and ITI. The ITI phase, which varied randomly between 4,000-6,000 ms, was treated as a rest phase against which key contrasts were run. Although incorporated into the design, response phase and motor response were modeled out of the analyses by assigning null weights to their regressors. The generating uses ITI contrast was

computed by assigning a weight of “1” to the regressor corresponding to the generation phase and a weight of “-1” to the regressor corresponding to ITI in the uses run (a weight of “0” was assigned to all other regressors). Across all participants, generating uses ITI activated the left dorsolateral PFC and anterior PFC (Table 1). Similarly, the *recalling* characteristics ITI contrast was computed by assigning a weight of “1” to the regressor corresponding to the characteristics phase and a weight of “-1” to the regressor corresponding to ITI phase in the characteristics run (a weight of “0” was assigned to all other regressors). Across all participants, recalling characteristics ITI activated the left dorsolateral PFC and right secondary somatosensory cortex (Table 1). The uses parameter was computed by assigning a weight of “1” to its corresponding regressor and a weight of “0” to all other regressors. This analysis activated right orbitofrontal cortex (Table 1). Similarly, the characteristics parameter was computed by assigning a weight of “1” to its corresponding regressor and a weight of “0” to all other regressors. This analysis activated the left supplementary motor cortex (Table 1). We also ran the generating uses–recalling characteristics contrast by assigning a weight of “1” to the regressor corresponding to the generating uses phase and a weight of “-1” to the regressor corresponding to the recalling characteristics phase (a weight of “0” was assigned to all other regressors). This contrast did not reveal any area of activation that survived our statistical threshold.

An independent-groups *t* test revealed that generating uses (versus ITI) resulted in significantly lower activation in right ventrolateral PFC (BA 47) in the experimental than active control group ($p = .024$) (Figure 3 and Table 2). The reverse contrast revealed that no region

was activated more in the experimental than active control group. We also compared the generating uses–recalling characteristics contrast between the experimental and active control groups using an independent-groups *t* test. This contrast did not reveal any area of activation that survived our statistical threshold.

Given that training (experimental vs. active control) was associated with variation in (a) fluid intelligence and (b) activation in right ventrolateral PFC, we tested the hypothesis that gain in fluid intelligence mediated the link between training and activation in right ventrolateral PFC. As shown in Figure 4 and consistent with behavioural results presented earlier, training (experimental vs. active control) predicted gain in fluid intelligence (i.e., post – pre RAPM scores) such that participants in the experimental condition had greater gain in fluid intelligence than participants in the active control condition. In addition, gain in fluid intelligence predicted activation in right ventrolateral PFC such that greater gain in fluid intelligence was related to lower activation in right ventrolateral PFC. For each participant, activation in right ventrolateral PFC was calculated by creating a sphere with a 15 mm radius around the coordinate that was activated at the group level ($x = 40, y = 24, z = -24$), and extracting the *T*-score corresponding to the local maximum. Critically, the effect of training on activation in right ventrolateral PFC through the mediator (i.e., fluid intelligence) was significant, Sobel-test $z = 1.99, p = .046$ (Preacher and Hayes, 2004).

In addition, previously Goel and Vartanian (2005) had shown that activation in right dorsolateral PFC (BA 46) covaries as a function of the *number* of generated solutions in a divergent thinking task. We therefore conducted another independent-groups *t* test based on

the parameters relating brain activation to the number of generated solutions (i.e., a parametric analysis). However, because of our interest in right dorsolateral PFC, we used Small Volume Correction in SPM8 to conduct a region-of-interest (ROI) analysis by creating a sphere with a 15 mm radius around the maximum activated in Goel and Vartanian (2005) ($x = 48, y = 38, z = 24$). The result revealed significantly lower activation in right dorsolateral PFC (BA 46) in the experimental than active control group in relation to the number of generated uses, corrected for multiple comparisons using the Bonferroni family-wise correction within the ROI ($p = .007$) (Figure 3 and Table 2). The reverse contrast revealed that no region was activated more in the experimental than active control group.

No significant differences were observed between the experimental and active control groups for recalling characteristics.

4 Discussion

We set out to test four hypotheses. First, we predicted an increase in WM capacity in participants enrolled in the experimental condition who trained on the n-back task. This prediction was confirmed as participants in the experimental group who trained on the WM task exhibited 12% improvement in performance in the WM task between the first and third sessions. Second, compared to participants enrolled in the active control condition, we predicted that participants enrolled in the experimental condition would exhibit improvement in fluid intelligence. This prediction was confirmed as unlike participants in the active control condition who exhibited no change in fluid intelligence, participants in the experimental condition exhibited an 8% improvement in fluid intelligence at the end of training compared to baseline. Third, compared to participants enrolled in an active control condition, we predicted

better performance on the AUT for participants in the experimental condition. This prediction was not supported. Fourth, we expected to observe variation in the BOLD response in the ventral and/or dorsal PFC as a function of WM training. This prediction was supported as activation in ventrolateral PFC (BA 47) was significantly lower during the generation phase of divergent thinking, as was activation in dorsolateral PFC (BA 46) in relation to the number of generated solutions in the task. Furthermore, although not hypothesized a priori, we also found that the effect of training on activation in right ventrolateral PFC was mediated by gain in fluid intelligence. This finding sheds light on the mechanism relating training to divergent thinking. These results demonstrate that although small gains made in WM capacity and fluid intelligence following a short and concentrated regimen of WM training did not transfer to better performance in divergent thinking, participants who underwent training exhibited neural efficiency during engagement in divergent thinking, and that this effect was mediated by gains in fluid intelligence. The extent to which the observed effect on neural efficiency is moderated by sex is an important question (see Neubauer et al., 2005), although it could not be addressed here because of the discrepancy between males and females recruited for the study. Furthermore, given that the generating uses–recalling characteristics contrast between the experimental and active control groups did not reveal any area of activation suggests that the effect of WM training on neural function was not specific to AUT.

Particularly interesting is the specificity of the observed reductions in activation in the ventrolateral (BA 47) and dorsolateral (BA 46) PFC exclusively. Goel and Vartanian (2005) investigated the generation of solutions using the Matchstick Task—another well-known measure used in the divergent thinking literature (Guilford, 1967). They reported activation in

right ventrolateral PFC ($x = 34, y = 16, z = -16$) when participants solved matchstick problems successfully, and in right dorsolateral PFC ($x = 48, y = 38, z = 24$) in relation to the number of generated solutions. The results of the present study indicate that WM training was associated with lower activation specifically in the two regions that are implicated in two dissociable aspects of divergent thinking. Because ventral and dorsal PFC are key parts of the neural correlates of WM, fluid intelligence, and divergent thinking, the specificity of the observed activations suggests that the ventral and dorsal aspects of PFC are the likely loci where transfer effects between WM training and divergent thinking could be observed.

An important consideration for interpreting the results of the present experiment involves the features of the response collected on each trial of the generation phase—a single digit indicating the number of generated uses (i.e., fluency). Although fluency scores account for the largest portion of the variance in divergent thinking tests (Plucker and Renzulli, 1999), they should be treated as one of several possible proxy measures of divergent thinking rather than a single measure of creative idea generation ability. This is especially true in light of recent evidence showing that whereas earlier responses in divergent thinking tasks tend to be retrieved from memory and are less likely to be creative, later responses tap executive functions and are more likely to be creative (Beaty and Silvia, in press; Gilhooly et al., 2007). Obtaining a single response that represents a combination of early and late responses does not allow one to distinguish between more and less creative uses in response to the same prompt. Although obtaining verbal or written responses in the fMRI scanner is technologically challenging, promising new devices should enable researchers to obtain responses that can later be scored for various outcome measures of interest beyond fluency, such as originality,

flexibility, elaboration, and creativity (Tam et al., 2010). These innovations will also enable researchers to score responses in novel ways (Silvia et al., 2008).

Furthermore, a response on any given trial represents the outcome of at least three different co-occurring processes. First, the participant must maintain in WM each generated use. Second, the participant must keep track of successful and unsuccessful attempts so that only instances of the former are added to the to-be-reported total at the end of each trial. Finally, there is a burden on conflict detection and/or resolution abilities because as each additional use is generated, there is a requirement to resolve the conflict between the earlier total that is no longer correct and the new updated response. There is ample evidence demonstrating that the PFC is involved in all aforementioned processes, namely maintenance functions in WM, monitoring of sub-goal processes, and conflict detection and/or resolution (see Goel and Vartanian, 2005). Therefore, the observed activations reflect not only the core processes related to divergent thinking, but also co-occurring “support” processes. However, these processes likely also affect responding in the characteristics phase, although perhaps to a different extent than they do in the generation phase.

4.1 Limitations

The present experiment had a number of limitations. First and foremost, because we did not collect pre-training behavioural and neural data for AUT, one cannot argue that WM training *causes* increased neural efficiency in divergent thinking. Using the logic of a between-subjects design (with respect to AUT), the present results warrant further investigation to determine the extent to which the variance observed in neural efficiency can be attributed to the treatment effect using a within-subjects design (with respect to AUT) in future studies.

Second, there is tremendous diversity in the tasks that have been used in active control conditions in previous WM training studies (see Buschkuhl and Jaeggi; 2010, Buschkuhl et al., 2012; Klingberg, 2010; Morrison and Chein, 2011). We selected our task specifically because (a) it is not considered a WM task, (b) its scores are not predictive of fluid intelligence, and (c) it is cognitively minimally taxing. However, arguably an active control task should have the exact same attributes as the WM training task, with the only exception being demand on WM. It is therefore important to remember that the task used in the present experiment in the active control condition differed from the n-back task used in the experimental condition in other attributes than demand on WM exclusively.

Third, following Jaeggi et al. (2008), we administered the RAPM within a time limit (10 minutes). Scores derived using this timed procedure have been criticized by some as an incorrect representation of fluid intelligence (Moody, 2009; but see Raven et al., 1998, Sternberg, 2008). It is possible that the gain observed in fluid intelligence as a function of WM training was due to the specific way in which RAPM was administered and scored, although it is important to keep in mind that gains in fluid intelligence as a function of WM training have been observed using a diversity of measures, and administration and scoring procedures in previous studies (see Buschkuhl et al., 2012; Klingberg, 2010; Morrison and Chein, 2011).

Fourth, although our results demonstrated an increase in performance on the n-back task from the first to the third session (Figure 1), it is arguable that part of this improvement might have been due to increased familiarity with the n-back task, rather than improvement in WM per se. This concern in interpreting WM training effects is not unique to this setting (see Klingberg, 2010), and addressing it will require assessments of how changes in performance in

the trained task correlate with measures of WM capacity and/or span at each training session. Particularly useful in contributing to this effort are studies on the correlation between the tasks used to implement WM training and measures of WM span (Jaeggi et al., 2010).

4.2 Implications

Our results should be considered within the contemporary view of intelligence as an emergent product of functional brain networks (Langer et al., 2012). This view is consistent with the two leading theoretical frameworks of the neuroscience of intelligence: The neural efficiency hypothesis, according to which there is an inverse relation between psychometric intelligence and brain metabolism (Neubauer and Fink, 2009), and the Parieto-Frontal Integration Theory (P-FIT), according to which intelligence is instantiated within a distributed parietal and frontal network (Jung and Haier, 2007), and correlated with brain activation during tasks varying in their loadings on general intelligence (Haier et al., 2003). Indeed, our result show that the effect of training on brain function is evident within the P-FIT network (i.e., VLPFC and DLPFC), consistent in direction with the neural efficiency hypothesis. In an exciting new contribution to the literature, Jung and Haier (in press) have recently proposed a Frontal Disinhibition Model (F-DIM) for creativity, which is neuroanatomically largely dissociable from P-FIT. We expect the examination of the interplay between P-FIT and F-DIM to be a major contributor to research on the neuroscience of creativity in the years ahead.

In terms of the broader cognitive training literature, the results are also relevant for the debate about the relation between duration of WM training and the direction of the observed neural effects (Buschkuhl et al., 2012; Klingberg, 2010). In the present study a relatively short duration of training was associated with a lower BOLD signal post training. However, the lower

BOLD signal is difficult to explain in terms of strategy learning and priming during encoding, because the training task (i.e., n-back) was different than the target task (i.e., AUT). It is also difficult to explain using time-on-task effects, because our design included an active control condition in which participants engaged in a different task but for an identical duration. In fact, our results demonstrated that the effect of training on activation in PFC was mediated by gain in fluid intelligence. This is an important new finding that sheds light on the possible mechanism relating training to divergent thinking. It is also possible that the correlation between neural efficiency and cognitive training was a function of the degree of overlap (i.e., specificity) of trained and target neural regions. This causal hypothesis remains to be tested in future studies.

References

- Baddeley AD (2003), Working memory: Looking back and looking forward. *Nat Rev Neurosci* 4:829-839.
- Beaty RE, Silvia PJ (in press), Why do ideas get more creative across time? An executive interpretation of the serial order effect in divergent thinking tasks. *Psychology of Aesthetics, Creativity, and the Arts*.
- Buschkuehl M, Jaeggi SM (2010), Improving intelligence: A literature review. *Swiss Med Wkly* 140:266-272.
- Buschkuehl M, Jaeggi SM, Jonides J (2012), Neuronal effects following working memory training. *Dev Cogn Neurosci* 2S:S167-S179.
- Cattell RB (1963), Theory of crystallized and fluid intelligence: A critical experiment. *J Educ Psychol* 54:1-22.
- Cohen J (1988), *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Deary IJ, Penke L, Johnson W (2010), The neuroscience of human intelligence differences. *Nat Rev Neurosci* 11:201-211.
- Duncan J, Seitz RJ, Kolodny J, Bor D, Herzog H, Ahmed A, Newell FN, Emslie H (2000), A neural basis for general intelligence. *Science* 289:457-460.
- Fink A, Grabner RH, Benedek M, Reishofer G, Hauswirth V, Fally M, Neuper C, Ebner F, Neubauer AC (2009), The creative brain: Investigation of brain activity during creative problem solving by means of EEG and fMRI. *Hum Brain Mapp* 30:734-748.
- Gilhooly KJ, Fiortou E, Anthony SH, Wynn V (2007), Divergent thinking: Strategies and executive

- involvement in generating novel uses for familiar objects. *Brit J Psychol* 98:611-625.
- Goel V, Vartanian O. (2005), Dissociating the roles of right ventral lateral and dorsal lateral prefrontal cortex in generation and maintenance of hypotheses in set-shift problems. *Cereb Cortex* 15:1170-1177.
- Gottfredson LS (1997), Why g matters: The complexity of everyday life. *Intelligence* 24:79-132.
- Gray JR, Chabris CF, Braver TS (2003), Neural mechanisms of general fluid intelligence. *Nat Neurosci* 6:316-322.
- Grushcow M (2008), Cognitive Test Software (Version 2.0) [Computer software]. North York, ON: NTT Systems Inc.
- Guilford JP (1967), *The nature of human intelligence*. New York: McGraw-Hill.
- Haier RJ, White NS, Alkire MT (2003), Individual differences in general intelligence correlate with brain function during nonreasoning tasks. *Intelligence* 31:429-441.
- Jaeggi SM, Buschkuhl M, Jonides J, Perrig WJ (2008), Improving fluid intelligence with training on working memory. *P Natl Acad Sci USA* 105:6829-6833.
- Jaeggi SM, Studer-Luethi B, Buschkuhl M, Su Y-F, Jonides J, Perrig WJ (2010), The relationship between n-back performance and matrix reasoning—Implications for training and transfer. *Intelligence* 38:625-635.
- Jung RE, Haier RJ (2007), A Parieto-Frontal Integration Theory (P-FIT) of intelligence: Converging neuroimaging evidence. *Behav Brain Sci* 30:135-187.
- Jung RE, Haier RJ (in press), Creativity and intelligence: Brain networks that link and differentiate the expression of genius. In: *Neuroscience of creativity* (Vartanian O, Bristol AS, Kaufman JC eds). Cambridge, MA: The MIT Press.

- Klingberg T (2010), Training and plasticity of working memory. *Trends Cogn Sci* 14:317-324.
- Langer N, Pedroni A, Gianotti LRR, Hänggi J, Knoch D, Jäncke L (2012), Functional brain network efficiency predicts intelligence. *Hum Brain Mapp* 33:1393-1406.
- Lee KH, Choi YY, Gray JR, Cho SH, Chae J-H, Lee S, Kim K (2006), Neural correlates of superior intelligence: Stronger recruitment of posterior parietal cortex. *Neuroimage* 29:578-586.
- Luders E, Narr KL, Thompson PM, Toga AW (2009), Neuroanatomical correlates of intelligence. *Intelligence* 37:156-163.
- Moody DE (2009), Can intelligence be increased by training on a task of working memory? *Intelligence* 37:327-328.
- Morrison A, Chein J (2011), Does working memory training work? The promise and challenges of enhancing cognition by training working memory. *Psychon B Rev* 18:46-60.
- Nakashima A, Vartanian O, Bouak F, Hofer K, Cheung B (2011), A test battery for the assessment of psychological and physiological performance following primary blast wave exposure. In: *Proceedings of the Military and Veterans Health Research Forum* (Aiken AB, Bélanger SAH, eds), pp 134-153. Kingston, ON: Canadian Defence Academy Press.
- Neubauer AC, Fink A (2009), Intelligence and neural efficiency. *Neurosci Biobehav R* 33:1004-1023.
- Neubauer AC, Fink A, Schrausser DG (2002), Intelligence and neural efficiency: The influence of task content and sex on the brain-IQ relationship. *Intelligence* 30:515-536.
- Neubauer AC, Grabner RH, Fink A, Neuper C (2005). Intelligence and neural efficiency: Further evidence of the influence of task content and sex on the brain-IQ relationship. *Cognitive Brain Res* 25:217-225.

- Nusbaum EC, Silvia PJ (2011), Are intelligence and creativity really so different?: Fluid intelligence, executive processes, and strategy use in divergent thinking. *Intelligence* 39:36-45.
- Oldfield RC (1971), The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia* 9:97-113.
- Olesen P, Westerberg H, Klingberg T (2004), Increased prefrontal and parietal brain activity after training of working memory. *Nat Neurosci* 7:75-79.
- Plucker JA (1999), Is the proof in the pudding? Reanalyses of Torrance's (1958 to present) longitudinal data. *Creativity Res J* 12:103-114.
- Plucker JA, Renzulli JS (1999), Psychometric approaches to the study of human creativity. In: *Handbook of creativity* (Sternberg RJ, ed.), pp 35-61. New York: Cambridge University Press.
- Prabhakaran V, Narayanan K, Zhao Z, Gabrieli JD (2000), Integration of diverse information in working memory within the frontal lobe. *Nat Neurosci* 3:85-90.
- Preacher KJ, Hayes AF (2004), SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behav Res Meth Ins C*, 36:717-731.
- Raven J, Raven JC, Court JH (1998), Raven manual: Section 4. Advanced progressive matrices. Oxford: Oxford Psychologists Press.
- Runco MA (1999), Divergent thinking. In: *Encyclopedia of creativity*, 2nd Ed. (Runco M, Pritzker S, eds.), pp 577-582. San Diego, CA: Academic Press.
- Silvia PJ, Winterstein BP, Willse JT, Barona CM, Cram JT, Hess KI, Martinez JL, Richard CA (2008), Assessing creativity with divergent thinking tasks: Exploring the reliability and validity of

- new subjective scoring methods. *Psychology of Aesthetics, Creativity, and the Arts*, 2:68-85.
- Sligh AC, Connors FA, Roskos-Ewoldsen B (2005), Relation of creativity to fluid and crystallized intelligence. *J Creative Behav* 39:123-136.
- Tam F, Churchill NW, Strother SC, Graham SJ (2010), A new tablet for writing and drawing during functional mri. *Hum Brain Mapp* 32:240-248.
- Unsworth N, Engle RW (2005), Working memory capacity and fluid abilities: Examining the correlation between Operation Span and Raven. *Intelligence* 33:67-81.
- Van der Heuvel MP, Stam CJ, Kahn RS, Hulshoff Poll, HC (2009), Efficiency of functional brain networks and intellectual performance. *J Neurosci* 29:7619 -7624.
- Vartanian O (2009), Variable attention facilitates creative problem solving. *Psychology of Aesthetics, Creativity, and the Arts* 3:57-59.
- Vartanian O (2011), Brain and neuropsychology. In: *Encyclopedia of creativity*, 2nd Ed., (Runco MA, Pritzker S, eds.), pp 164-169. San Diego, CA: Academic Press.
- Wiley J, Jarosz AF (2012), Working memory capacity, attentional focus, and problem solving. *Curr Dir Psychol Sci* 21:258-262

Table 1. Main effects for generating uses and recalling characteristics in the Alternate Uses

Task.

Task & Contrast	Structure	BA	x	y	z	T-score	k
Generating uses ITI	DLPFC	46/45	50	28	14	14.85	8297
	Anterior PFC	10	10	66	22	6.08	9
Uses parameter	OFC	10	4	44	2	10.40	3553
Recalling characteristics ITI	DLPFC	46/45	48	20	26	9.72	4730
	SII	1	50	32	6	5.53	375
Characteristics parameter	SMA	6	2	28	44	7.91	5326

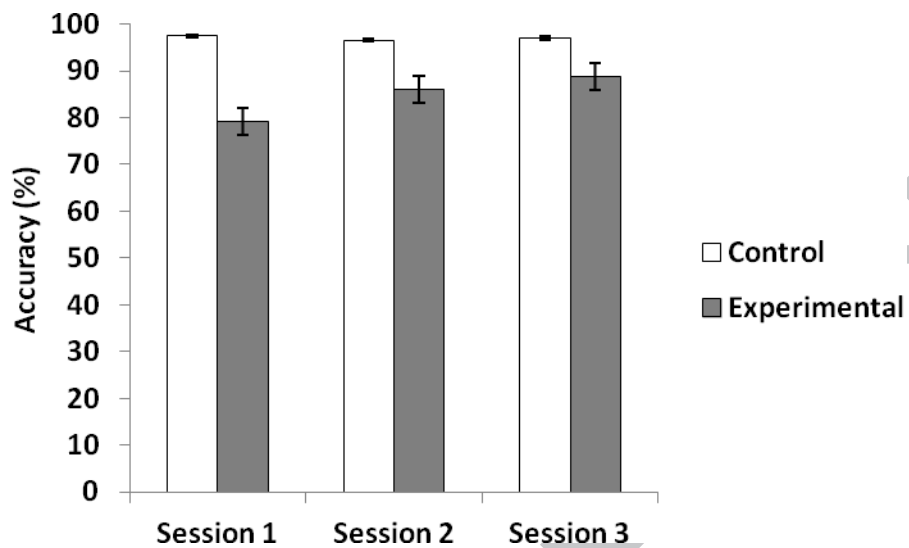
Notes. BA = Brodmann Area, k = cluster size (number of contiguous voxels), ITI = inter-trial interval (rest), DLPFC = dorsolateral prefrontal cortex, OFC = orbitofrontal cortex, SMA = supplementary motor cortex, SII = secondary somatosensory cortex. The coordinates are reported in MNI space. These structures represent only a subset of structures that survived a voxel-level intensity threshold of $p < .05$, corrected for multiple comparisons using the Bonferroni family-wise (whole-brain) correction (see *Method*). Each reported contrast activated other structures that did not survive this strict cut-off.

Table 2. Areas of significantly lower activation in the trained group in the Alternate Uses Task.

Contrast	Structure	BA	x	y	z	T-score	k
Generating uses	VLPFC	47	40	24	24	5.86	195
Number of uses	DLPFC	46	52	42	10	4.41	15

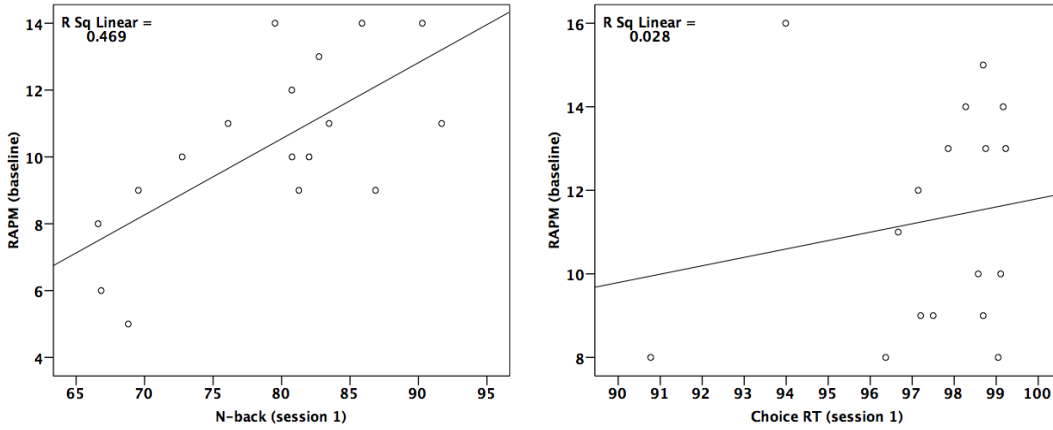
Notes. BA = Brodmann Area, *k* = cluster size (number of contiguous voxels), VLPFC = ventrolateral prefrontal cortex, DLPFC = dorsolateral prefrontal cortex. The coordinates are reported in MNI space. These structures represent only a subset of structures that survived a voxel-level intensity threshold of $p < .05$, corrected for multiple comparisons using the Bonferroni family-wise (whole-brain) correction (see *Method*). The activation in DLPFC involves ROI analysis (see *Results*). Each reported contrast activated other structures that did not survive this strict cut-off.

Figure 1. Impact of training on performance (% accuracy) across three training sessions.



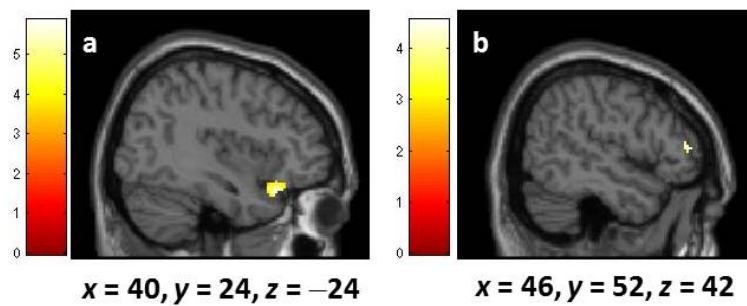
Notes. For the experimental condition, the grey bars represent accuracy (%) on the n-back task. For the active control condition, the white bars represent accuracy (%) on the choice RT task. Error bars represent standard error. All three levels corresponding to the experimental condition are significantly different from one another, whereas none of the levels corresponding to the active control condition differ from one another (see *Results*).

Figure 2. Performance on n-back (experimental task) but not 4-choice RT (active control task) predicted baseline RAPM scores.



Notes. In the Choice RT figure, the removal of the two outliers (to the left) increased R^2 Linear to 0.114; however, choice RT scores did not predict RAPM scores, $r = .34$, $p = .218$.

Figure 3. Impact of WM training on neural activation in divergent thinking.



Notes. (a) There was lower activation in right ventrolateral prefrontal cortex (BA 47) in the experimental group when generating uses in the Alternate Uses Task; (b) there was lower activation in right dorsolateral prefrontal cortex (BA 46) in the experimental group in relation to the number of generated solutions in the Alternate Uses Task based on an ROI analysis (see *Results*). SPM rendered into standard stereotactic space and superimposed on to sagittal MRI in standard space. Each bar represents the corresponding *T*-score.

Figure 4. Beta values for the mediator model of the effect of training on brain activation in right ventrolateral PFC (standard errors in parentheses).



Notes. VLPFC = Ventrolateral prefrontal cortex. * $p = .021$, ** $p = .002$.

Highlights of *Working Memory Training Is Associated with Lower Prefrontal Cortex*

Activation in a Divergent Thinking Task:

- We examined the effects of working memory (WM) training on divergent thinking.
- WM training led to improvements in WM capacity and fluid intelligence.
- WM training did not improve divergent thinking performance.
- WM training was correlated with lower prefrontal activation.
- Gain in fluid intelligence mediated the effect of training on activation in prefrontal cortex.