On the Malleability of Human Cognition: Working Memory Training and Transfer

Clark, Cameron

doctoral thesis

University of Calgary graduate students retain copyright ownership and moral rights for their thesis. You may use this material in any way that is permitted by the Copyright Act or through licensing that has been assigned to the document. For uses that are not allowable under copyright legislation or licensing, you are required to seek permission.

Downloaded from PRISM: https://prism.ucalgary.ca
UNIVERSITY OF CALGARY

On the Malleability of Human Cognition: Working Memory Training and Transfer

by

Cameron Mackenzie Clark

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE

DEGREE OF DOCTOR OF PHILOSOPHY

GRADUATE PROGRAM IN CLINICAL PSYCHOLOGY

CALGARY, ALBERTA

JANUARY, 2017

© Cameron Mackenzie Clark 2017
Abstract

Training working memory (WM) to increase WM capacity and fluid intelligence (Gf) has received much experimental attention in recent years, though its efficacy remains highly controversial. The current study investigated the effect of a randomized six-week online WM intervention on cognitive abilities and patterns of neural activation in a community-recruited sample of healthy young adults, in relation to both a processing speed training active control condition, as well as a no-contact control condition. Results of this randomized trial are discussed in three parts:

Chapter 2 examines group-level fMRI activation patterns for tasks of WM and Gf before the training intervention. Consistent with previous research, results indicate large areas of fronto-parietal activation in response to increasing task demands for our WM task, which largely subsume more circumscribed regions of activation for our Gf task. These results are discussed in terms of a task-general central network which may underlie performance of WM, Gf, and perhaps even goal-directed behaviour more generally.

Chapter 3 investigates potential differences in a wide range of cognitive test scores before and after WM training, processing speed training, or no-contact. Results revealed support for the null hypothesis across all cognitive tests administered. Because these results are consistent with experimental trials of equal or greater methodological rigor, we suggest that future research re-focus on promising interventions known to increase memory performance in healthy young adults; and/or examine alternative populations in which WM training may be efficacious.

Chapter 4 examines potential differences in pre- and post-training patterns of neural activation for WM and Gf tasks in our WM training, and processing speed training groups. Results indicated significant post-training reductions in activation for the WM trained group in
relation to the processing speed group for the WM task, but not the Gf task. These results suggest that WM training does not affect patterns of neural activation for Gf tasks. We suggest that future research investigate neural correlates of WM training in populations for which WM itself is impaired; and/or WM training interventions in populations that have returned more promising results compared to those with healthy young adults.

*Keywords:* cognitive training, working memory, working memory training, fluid intelligence, intelligence, reasoning, near-transfer, far-transfer, fMRI
Acknowledgements

I would like to thank my research supervisor, Dr. Vina Goghari, for her support throughout this dissertation process. I would also like to thank the members of my supervisory committee, Dr. Stewart Longman and Dr. Bradley Goodyear, for their invaluable contributions to design, execution, and analysis of the project and its resulting data. Additionally, I would also like to thank Dr. Tavis Campbell for assuming supervisory responsibilities in Dr. Goghari’s absence, and Drs. Ramasubbu and Yang for their contributions to my supervisory committee.

Thank you to all the participants who dedicated their time and efforts to this project. Six weeks of dedicated cognitive training is a big ask, and I and the research team were consistently amazed with the dedication and commitment our participants brought to the task. Without the help, support, and guidance of all members of the Seaman Family MR Centre, none of this project’s MRI data could have been collected. Special thanks to Frances Raymond for her consistent dedication to participant scheduling, and Filomeno Cortese, Dan Pittman, and Jolyn D’Andrea for their dedication to extracting only the highest quality data from our participants – and also for passing the scanning time with conversation perhaps even more interesting than seeing inside the heads of living human beings. Similarly, without the help of my labmate and co-investigator Linette Lawlor-Savage, and the team of spectacular research assistants she assembled, none of the behavioural data would have been collected either, much less organized. Thanks to Aiko Dolatre, Averi House, Emma Harris, and Naomi Rose-Dutta for their dedication to helping with the project.

Thanks also to Lumos Labs who were generous in their provision of the online training programs utilized by this project. Given the controversial nature of “brain training” I highly
commend Lumos’ open approach to collaborative research on the topic with their innovative technologies.

I am extremely grateful for all of the financial support that I and this project have received from a variety of private, provincial, and national sources, including: the T. Chen Fong Doctoral Research Excellence Scholarship in Medical Imaging Science, a graduate studentship from Alberta Innovates Health Solutions, the Vanier Canada Graduate Scholarship, as well as funds from the National Sciences and Engineering Research Council of Canada awarded to Dr. Goghari for her broader program of research.

I owe all of my successes to the incredible people I’ve had the benefit of knowing and working with at the University of Calgary and beyond. To my labmates, Briana Cassetta, Linette Lawlor-Savage, and Michael Spilka who have been more than just ‘other humans in the same room with computers’ – you are the best teammates a graduate student could hope for. To my cohort, Jennifer Ference, Melanie Khu, Leanne Quigley, and Igor Yakovenko – it truly has been a special fortune to embark on this educational journey with you, and seek answers to the ubiquitous question “What are the chances?” To my favourite colleague – you have enlivened and impassioned this journey beyond all reasonable prospect, and have made me feel lucky more than I had any right to expect. Finally, to my parents, Lois and Brian, your unwavering and unconditional love and support over the years simply bankrupts my ability to express my gratitude in words.

Thank you all.
# Table of Contents

Abstract.............................................................................................................................. ii
Acknowledgements............................................................................................................ iv
Table of Contents............................................................................................................... vi
List of Figures ...................................................................................................................... x
List of Abbreviations .......................................................................................................... xiii

## CHAPTER ONE: INTRODUCTION..................................................................................1

## CHAPTER TWO: COMPARING BRAIN ACTIVATIONS ASSOCIATED WITH WORKING MEMORY AND FLUID INTELLIGENCE .............................................11

2.1 Background................................................................................................................... 11
   2.1.1 Neural Mechanisms Associated with the n-Back Task ........................................ 13
   2.1.2 Neural Mechanisms Associated with Raven’s Progressive Matrices ..................... 16
   2.1.3 Neural Mechanisms Associated with the Lexical Decision Task ............................. 17
   2.1.4 The Neuroanatomy of Intelligence ..................................................................... 18
2.2 Method .......................................................................................................................... 20
   2.2.1 Participants and Recruitment .............................................................................. 20
   2.2.2 Procedure and Materials .................................................................................... 21
      2.2.2.1 Dual n-Back Functional Imaging Task .......................................................... 22
      2.2.2.2 Raven’s Progressive Matrices Functional Imaging Task ................................ 23
      2.2.2.3 Lexical Decision Functional Imaging Task .................................................... 23
   2.2.3 MRI Data Acquisition .......................................................................................... 24
   2.2.4 Data Analysis ...................................................................................................... 25
2.3 Results ........................................................................................................................... 26
   2.3.1 Participant Demographics and Cognitive Characteristics ..................................... 26
   2.3.2 In-Scanner Behavioural Data ............................................................................... 26
   2.3.3 Whole Brain fMRI Analyses .............................................................................. 29
      2.3.3.1 Dual n-Back Task, by Level of Difficulty ....................................................... 29
      2.3.3.2 Raven’s Progressive Matrices Task, by Level of Difficulty ............................. 34
      2.3.3.3 Lexical Decision Task, by Level of Difficulty .............................................. 36
   2.4 Discussion ................................................................................................................... 37
   2.5 References .................................................................................................................. 49

## CHAPTER THREE: WORKING MEMORY TRAINING AND TRANSFER TO UNTRAINED COGNITIVE ABILITIES: SUPPORT FOR THE NULL ......................59

3.1 Background ................................................................................................................... 59
3.2 Method ........................................................................................................................... 72
   3.2.1 Participants and Recruitment .............................................................................. 72
   3.2.2 Procedure and Materials .................................................................................... 74
   3.2.3 Cognitive Testing and Behavioural Measures ....................................................... 75
      3.2.3.1 Wechsler Adult Intelligence Scale – Fourth Edition (WAIS-IV) ..................... 76
      3.2.3.2 Raven’s Advanced Progressive Matrices (RAPM) ......................................... 77
      3.2.3.3 Cattell’s Culture Fair Test (CCFT) ................................................................. 77
      3.2.3.4 Automated Operation Span Task (AOSPAN) ................................................ 78
      3.2.3.5 Spatial Delayed Response Task (SDRT) ......................................................... 78
4.4.5 Conclusions .................................................................................................................. 184
4.5 References .................................................................................................................... 186

CHAPTER FIVE: GENERAL DISCUSSION .................................................................. 198
  5.1 Does Working Memory Training Work? .................................................................... 198
  5.2 How Reasonable are the Proposed Mechanisms for Transfer of WM Training?.. 202

REFERENCES ..................................................................................................................... 210
List of Tables

Table 1: Chapter 2: Significant whole-brain activations by task and contrast. .............................. 29

Table 2: Chapter 3: Test-retest reliability correlations for cognitive tests, based on scores from the no-contact control group. ................................................................. 78

Table 3: Chapter 3: Participant characteristics .................................................................................... 86

Table 4: Chapter 3: Bayesian factors for time × group interactions by measure resulting from JZS Bayesian repeated measures ANOVAs. ......................................................... 95

Table 5: Chapter 4: Participant characteristics. ..................................................................................... 146

Table 6: Chapter 4: Significant whole-brain activations by group, task, and contrast .......... 152

Table 7: Chapter 4: Significant whole-brain activations identified by significant time × group interaction terms ............................................................................................................. 156
List of Figures

Figure 1: Chapter 2: Performance and reaction time data for in-scanner tasks: (A) dual n-back task, (B) Raven’s Standard Progressive Matrices, and (C) a Lexical Decision Task. Error bars represent 95% confidence intervals........................................ 28

Figure 2: Chapter 2: Significant whole brain activation for contrast 3-back > 1-back of the Dual n-Back Task. ................................................................. 31

Figure 3: Chapter 2: Significant whole brain activation for contrast 2-back > 1-back of the Dual n-Back Task. ................................................................. 33

Figure 4: Chapter 2: Significant whole brain activation for contrast Hard > Medium of the Raven’s Standard Progressive Matrices Task................................................................. 34

Figure 5: Chapter 2: Significant whole brain activation for contrast Hard > Easy of the Raven’s Standard Progressive Matrices Task................................................................. 35

Figure 6: Chapter 2: Significant whole brain activation for contrast Hard > Easy of the Lexical Decision Task................................................................. 37

Figure 7: Chapter 2: Harder > Easier difficulty contrasts for all tasks overlaid. Panel A shows the overlap of the dual n-back task (blue) and Raven’s Standard Progressive Matrices (green); and panel B shows the overlap of the dual n-back task (blue) with the lexical decision task. Images displayed in radiological convention................................................................. 42

Figure 8: Chapter 3: Flow chart of study design. *Two participants in MRI conditions were reassigned to the non-contact control group after being unable to tolerate MRI scanning. †Participants removed from analysis due to training contamination, low training dosage, or data acquisition issues................................................................. 73

Figure 9: Chapter 3: Lumosity working memory training tasks (panel A) and processing speed active control tasks (panel B)................................................................. 82

Figure 10: Chapter 3: Mean performance by training game for the working memory training group (A-C), and mean reaction times by training game for the processing speed training group (D-F).................................................................................. 89

Figure 11: Chapter 3: WAIS-IV performance by group, before and after training. Verbal Comprehension Index (A); Perceptual Reasoning Index (B); Working Memory Index (C); Processing Speed Index (D); Full Scale Intelligence Quotient (E). Error bars represent 95% confidence intervals................................................................. 92

Figure 12: Chapter 3: Automated Operation Span (A), Spatial Delayed Response Task (B-C), Raven’s Advanced Progressive Matrices (D), and Cattell’s Culture Fair Test (E) performance by group before and after training. Error bars represent 95% confidence intervals................................................................................ 94
Figure 13: Chapter 3: Participant self-rated motivation to complete training (A), and expectations for improvement (B) on the training tasks throughout the training period. Error bars represent 95% confidence intervals.

Figure 14: Chapter 4: Flow chart of study design. *Two participants in MRI conditions were reassigned to the no-contact control group after being unable to tolerate MRI scanning. †Participants removed from analysis due to training contamination, low training dosage, or data acquisition issues.

Figure 15: Chapter 4: In-scanner cognitive tasks: 2-back condition of the nback task (A); adapted Raven’s Standard Progressive Matrices (B); and, the lexical decision task (C), with correct answers indicated.

Figure 16: Chapter 4: Lumosity working memory training tasks (panel A) and processing speed active control tasks (panel B).

Figure 17: Chapter 4: Mixed effects repeated measures ANOVA group level model examining time (within-subjects; pre- vs. post-training), and group (between subjects; working memory training vs. processing speed active control training).

Figure 18: Chapter 4: Mean performance by training game for the working memory training group (A-C), and mean reaction times by training game for the processing speed training group (D-F).

Figure 19: Chapter 4: In-scanner performance on the n-back task (A), and Raven’s Standard Progressive Matrices (B) by group, before and after training. Participants in the WM group performed better on the 3-back condition after training, and participants in the PS group performed worse on the hardest level of the Ravens Progressive Matrices task after training. Error bars represent 95% confidence intervals.

Figure 20: Chapter 4: Participants self-reported motivation to complete training (A), and self-reported expectations for cognitive improvement as a result of training (B) on the training tasks throughout the training period. Error bars represent 95% confidence intervals.

Figure 21: Chapter 4: Significant whole brain activations for the 3-back > 2-back contrast of the dual n-back task.

Figure 22: Chapter 4: (A) Significant whole-brain time × group interaction activations for the 3-back > 2-back contrast of the dual n-back task. (B) Mean signal intensity in contrast of parameter estimate (COPE) values by time and group for the voxels identified in panel A. Error bars represent 95% confidence intervals.

Figure 23: Chapter 4: Significant whole brain activations for the 3-back > 1-back contrast of the dual n-back task.
Figure 24: Chapter 4: (A) Significant whole-brain time × group interaction activations for the 3-back > 1-back contrast of the dual n-back task. (B) Mean signal intensity in contrast of parameter estimate (COPE) values by time and group for the voxels identified in panel A. Error bars represent 95% confidence intervals. ................................................. 163

Figure 25: Chapter 4: Significant whole brain activations for the 2-back > 1-back contrast of the dual n-back task. ................................................................. 165

Figure 26: Chapter 4: (A) Significant whole-brain time × group interaction activations for the 2-back > 1-back contrast of the dual n-back task. (B) Mean signal intensity in contrast of parameter estimate (COPE) values by time and group for the voxels identified in panel A. Error bars represent 95% confidence intervals. ................................................. 166

Figure 27: Chapter 4: Significant whole brain activations for the Hard > Medium contrast of the Raven’s Standard Progressive Matrices task. ......................................................... 168

Figure 28: Chapter 4: (A) Significant whole-brain time × group interaction activations for the Hard > Medium contrast of the Raven’s Progressive Matrices task. (B) Mean signal intensity in contrast of parameter estimate (COPE) values by time and group for the voxels identified in panel A. Error bars represent 95% confidence intervals................. 169

Figure 29: Chapter 4: Significant whole brain activations for the Hard > Easy contrast of the Raven’s Standard Progressive Matrices task. ................................................................. 171

Figure 30: Chapter 4: Significant whole brain activations for the Medium > Easy contrast of the Raven’s Standard Progressive Matrices task. ................................................................. 172

Figure 31: Chapter 4: Significant whole brain activations for the Hard > Easy contrast of the lexical decision task. ................................................................. 174
## List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>AOSPAN</td>
<td>Automated Operation Span Task</td>
</tr>
<tr>
<td>CCFT</td>
<td>Cattell’s Culture Fair Test</td>
</tr>
<tr>
<td>CHREB</td>
<td>Conjoint Health Research Ethics Board</td>
</tr>
<tr>
<td>COPE</td>
<td>Contrast of Parameter Estimate</td>
</tr>
<tr>
<td>DLPFC</td>
<td>Dorsolateral Prefrontal Cortex</td>
</tr>
<tr>
<td>EPI</td>
<td>Echo Planar Imaging</td>
</tr>
<tr>
<td>FEAT</td>
<td>fMRI Expert Analysis Tool</td>
</tr>
<tr>
<td>FLIRT</td>
<td>FMRIB Linear Image Registration Tool</td>
</tr>
<tr>
<td>fMRI</td>
<td>Functional Magnetic Resonance Imaging</td>
</tr>
<tr>
<td>FMRIB</td>
<td>Functional MRI of the Brain</td>
</tr>
<tr>
<td>FOV</td>
<td>Field of View</td>
</tr>
<tr>
<td>FSIQ</td>
<td>Full Scale Intelligence Quotient</td>
</tr>
<tr>
<td>FSL</td>
<td>FMRIB Software Library</td>
</tr>
<tr>
<td>FWHM</td>
<td>Full Width Half Maximum</td>
</tr>
<tr>
<td>Gc</td>
<td>Crystallized Intelligence</td>
</tr>
<tr>
<td>Gf</td>
<td>Fluid Intelligence</td>
</tr>
<tr>
<td>IQ</td>
<td>Intelligence Quotient</td>
</tr>
<tr>
<td>LDT</td>
<td>Lexical Decision Task</td>
</tr>
<tr>
<td>MP-RAGE</td>
<td>Magnetization Prepared Rapid Acquisition Gradient Echo</td>
</tr>
<tr>
<td>MRI</td>
<td>Magnetic Resonance Imaging</td>
</tr>
<tr>
<td>ms</td>
<td>Milliseconds</td>
</tr>
<tr>
<td>PET</td>
<td>Photon Emission Tomography</td>
</tr>
<tr>
<td>P-FIT</td>
<td>Parieto-Frontal Integration Theory of Intelligence</td>
</tr>
<tr>
<td>PRI</td>
<td>Perceptual Reasoning Index of the WAIS-IV</td>
</tr>
<tr>
<td>PSI</td>
<td>Processing Speed Index of the WAIS-IV</td>
</tr>
<tr>
<td>RAPM</td>
<td>Raven’s Advanced Progressive Matrices</td>
</tr>
<tr>
<td>RPM</td>
<td>Raven’s Progressive Matrices</td>
</tr>
<tr>
<td>RSPM</td>
<td>Raven’s Standard Progressive Matrices</td>
</tr>
<tr>
<td>SD</td>
<td>Standard Deviation</td>
</tr>
<tr>
<td>SDRT</td>
<td>Spatial Delayed Response Task</td>
</tr>
<tr>
<td>TE</td>
<td>Echo Time</td>
</tr>
<tr>
<td>TI</td>
<td>Inversion Time</td>
</tr>
<tr>
<td>TR</td>
<td>Repetition Time</td>
</tr>
<tr>
<td>VCI</td>
<td>Verbal Comprehension Index of the WAIS-IV</td>
</tr>
<tr>
<td>WAIS-IV</td>
<td>Wechsler Adult Intelligence Scale – Fourth Edition</td>
</tr>
<tr>
<td>WM</td>
<td>Working Memory</td>
</tr>
<tr>
<td>WMI</td>
<td>Working Memory Index of the WAIS-IV</td>
</tr>
</tbody>
</table>
Chapter One: Introduction

Intelligence is perhaps the most extensively studied and best understood concept in all of psychology. However, despite a history that dates back as long as experimental psychology itself, challenges to its definition and measurement have been many, and certainly persist to the present day (Gottfredson & Saklofske, 2009). Aimed at ameliorating misunderstandings of intelligence and its disparate definitions, a collaboration of 52 scientists and researchers produced the following consensus definition: “Intelligence is a very general mental capability that among other things, involves the ability to reason, plan, solve problems, think abstractly, comprehend complex ideas, learn quickly, and learn from experience” (Gottfredson, 1997, pp. 13; as cited in Nisbett et al. 2012, pp. 2). Intelligence, so defined and measured, holds massive practical importance as these skills have been found to be strongly associated with educational, occupational, economic, and social outcomes such as school attainment, job success, and even longevity (Gottfredson, 1997; Gottfredson & Saklofske, 2009). It follows from these empirical links that higher intelligence should confer an advantage in each of these broad domains, while lower intelligence should be conversely disadvantageous – though neither is a guarantor of success or failure in any domain.

Following this logic, much research has been devoted to the endeavor of increasing intelligence in the hopes of positively affecting these broad-based and strongly associated outcomes. Indeed, in his review of programs intended to raise intelligence, A.R. Jensen noted that “few other topics in the history of behavioural science have resulted in so vast literature, and no other comes close to it in total research expenditure” (Jensen, 1998, pp. 333). At the time of Jensen’s (1998) review, results had been disappointing. Small to moderate increases could be facilitated in specific skills, however observed gains faded quickly over time, and no intervention
seemed able to facilitate increases in a more general intelligence or ‘g’ as proposed by Spearman as early as 1904 (Neisser et al., 1996; Gottfredson, 1997; Jensen, 1998).

In fact, as far back as the 19th century, pioneers of empirical psychology were working diligently to discover whether or not practice with one task might generalize to increased performance on another. William James himself investigated this idea by analyzing the speed with which he could memorize poems. He reasoned that practice in memorization of one poem should improve the speed at which he could memorize another. After committing two large sections of different poems to memory, he found that the second poem had actually taken a greater amount of time per line to memorize (50 seconds versus 57 seconds; James, 1890). He subsequently repeated the experiment with four students and found equivocal results, with reductions in average per-line memorization time for only two of these four students. He concluded: “one’s native retentiveness is unchangeable” (pp. 663)

Several years later, Thorndike and Woodworth (1901a, 1901c, 1901c) described a series of similar investigations in a set of papers with titles that might not seem so out of place in scientific journals of intelligence and cognition today: “The influence of improvement in one mental function upon the efficiency of other functions”. Young adults trained in the estimation of the area of rectangles demonstrated increased performance for the task over the course of training, however their improved abilities for area estimation did not transfer to estimation of other object dimensions such as weight, nor even to area estimation for other shapes. Similarly, Thorndike and Woodworth found that participants trained to mark passages of text for words with certain characteristics (e.g. containing the letters ‘e’ and ‘s’) improved in speed and accuracy on this task, but showed no facilitation on either measure when asked to scan text for misspelled words, or even other letter combinations. Lastly, Thorndike (1923) tested the then-
popular notion that learning Latin ought to lead to widespread transfer of ability to various domains of cognition by promoting a ‘disciplined mind’. Similar to his previous findings, he found no evidence of widespread benefit from his participants’ linguistic erudition.

The idea that training of any kind might lead to generalized increases in untrained task ability through discipline or a disciplined mind predates even this early work by Thorndike. Catherine Aiken’s (1899) book *Exercises in mind-training in quickness of perception concentrated attention and memory*, clearly explained the rationale behind its prescribed series of mental drills (e.g. digit/letter recall, mental arithmetic, synonyms etc.): “The discipline derived from the daily practice of each group of exercises prepares the mind to seize upon the succeeding group, and gradually to acquire habits of alertness and attention” (pp. vi). Aiken even lays out the ostensibly modern logic of not simply teaching specific task-based skills, but rather general skills that ought to be useful in a variety of wider applications: “Let it be constantly borne in mind that the object of this book is not primarily to suggest methods of teaching, but methods of training the mind to receive instruction” (pp. vi; emphasis original). However, contrary to Thorndike’s empirical findings, Aiken did make the claim that her method produced the intended effects “…these methods have been tried and found distinctly efficacious as a means to an end” (pp. vi).

Despite the early empirical evidence against the idea of training transfer, psychological research throughout the 20th century continued to investigate the transferability of training across several psychological and educational disciplines. For example, perceptual learning research in the 1980s and 1990s found that practice with discrimination of vertical waveforms does not transfer to increased ability to discriminate horizontal waveforms (Fiorentini & Berardi, 1980), nor does learning to discriminate motion in one directional axis transfer to greater ability with
respect to different axes (Ball & Sekuler, 1982). Further, training in discrimination tasks may not transfer from one eye to the other (Fahle, Edelman, & Poggio, 1995), nor even to another location on the retina within the same visual field (Shiu & Pashler, 1992). In perhaps the most striking example of cognitive task improvement resulting from training, Ericsson, Chase, and Faloon (1980) reported an increase in memory span capacity in one undergraduate research volunteer from seven to 79 digits over the course of 230 hours of in-lab training. When switched to recall of letters versus digits however, his capacity dropped back to average recall performance of around six items. Research on higher-level cognitive processes has returned similar null results. Pea and Kurland (1984) found that general problem solving is not improved by learning computer programming (Pea & Kurland, 1984), though interestingly, their logic for why they thought it might have sounds rather familiar: “…programming may spontaneously discipline thinking.” (Pea, Kurlund, & Hawkins, 1985; pp. 196). Even more generally, Catrambone & Holyoak (1989) found that learning the solution to a given problem did not guarantee the application of that solution to similar problems presented in different contexts.

Taken together, these repeated examples of failure of transfer reaffirm Thorndike’s conclusions from nearly a century earlier: “the mind is so specialized into a multitude of independent capacities that we alter human nature only in small spots” (pp. 246). However, these studies in intelligence and transfer of task-based ability are critically important for two primary reasons. First, from an experimental perspective, they demonstrate the stubborn specificity of training and learning. That is, repetitive and intensive practice with almost any non-trivial task reliably leads to facilitation of performance on that task (though not so in the case of speeded poem memorization, apparently), but scientific investigations have consistently failed to find the transfer of that improved ability to almost any separate task. This holds even when the transfer
task is extremely similar to the trained task. Second, these early forays in cognitive training and transfer are important because they demonstrate not only the *prima facie* reasonability of the notion of transfer of training, but also the tenacity of the concept over time despite repeated empirical disconfirmation over the past century. In this sense, whether or not cognitive training has true effects on domain-general abilities or general intelligence more broadly, the fact that we believe it ought to says something very true about us. The reasonability of transfer persists in large part because it drives at a central notion of education itself, which is to teach skills and train aptitudes that are *broadly* useful in performing complex tasks and solving problems in the service of achieving larger goals. In Aiken’s (1899) language, we might say that the *means* have changed over time (i.e. different methods of training), but the *end*, or the desire for improved cognition and greater intelligence, certainly has not. Likewise, and even more specifically, the kinds of cognitive tasks that appear as though they ought to transfer to more general abilities, as well as the mechanisms by which they ought to, have changed remarkably little over the last century. Perhaps no greater proof of this point exists than the fact that at the time of this writing, Aiken’s 1899 book (though freely available online) is for sale on Amazon.com as an e-book under the new title *Mind Training Games and Exercises* – published in 2013, 114 years after its first publication. Anecdotally, it warrants comment that the sole customer review that this e-book received pertains to font size and page formatting rather than the suitability of the prescribed training tasks, or their efficacy. Certainly, the opinions of any one customer cannot be taken as representative of the beliefs of society at large, though it is difficult to imagine any other variety of century-old scientific idea masquerading as current to such a degree.

This historical overview provides the context necessary to appreciate the burgeoning and controversial literature on working memory (WM) training. As only the latest in a long line of
empirical investigations examining transfer effects, researchers are both rightly excited at the prospect of raising domain-general abilities because they have long been thought to be immutable – and rightly skeptical for the very same reason (e.g. Sternberg, 2008; Slagter, 2012). Relatively recent investigations have found WM training to be effective in significantly raising measured fluid intelligence (Jaeggi, Buschkuehl, Jonides, & Perrig, 2008; Jaeggi, Studer-Luethi, Buschkuehl, Su, Jonides & Perrig, 2010; Jaeggi, Buschkuehl, Jonides & Shah, 2011), and support the notion the adult brain is capable of much greater neurocognitive plasticity than previously thought. Unfortunately however, less than ideal research methodology at the level of individual studies (see Moody, 2009; Button et al, 2013; Boot, Simons, Stothart, & Stutts, 2013; Redick, 2015; Melby-Lervåg, Redick, & Hulme, 2016; Moreau, Kirk, & Waldie, 2016), inconsistent focus on populations of interest (e.g. age, clinical status), lack of theory-driven approaches to task and measure selection (von Bastian & Oberauer, 2013), and inconsistent study inclusion criteria at the level of meta-analysis (Melby-Lervåg & Hulme, 2013; Au, Sheehan, Tsai, Duncan, Buschkuehl, & Jaeggi, 2014; Melby-Lervåg et al., 2016) have jointly contributed to a largely polarized literature on the efficacy of WM training.

Add to this the issue that many ‘brain training’ programs have been proffered and vaunted by for-profit companies with widespread marketing campaigns extolling their efficacy, and the academic question over cognitive training expands to the realms of public health policy, and business ethics. As an example, Redick and colleagues (2013) quote a statement taken directly from Lumosity’s website (date unknown): “Based on extensive research, Lumosity improves memory, attention, processing speed, and problem-solving skills so you can feel more confident in your abilities” (pp. 359). Industry claims such as this prompted the back and forth nature of the academic debate on WM training to spill out of academic journals into the public
domain. In 2014, a group of 60-70 doctors and scientists spoke out in a consensus statement on “brain games”; in which they stated:

…at this point it is not appropriate to conclude that training-induced changes go significantly beyond the learned skills, that they affect broad abilities with real-world relevance, or that they generally promote ‘brain health’. (http://longevity3.stanford.edu/blog/2014/10/15/the-consensus-on-the-brain-training-industry-from-the-scientific-community/)

However, in an almost perfect expression of the divided nature of the science, a parallel and opposing consensus statement was drafted and posted in response by a separate group of researchers, which stated the case for optimism with respect to WM training:

Many…studies show improvements that encompass a broad array of cognitive and everyday activities, show gains that persist for a reasonable amount of time, document positive changes in real-life indices of cognitive health, and employ control strategies designed to account for “placebo” effects. While we can debate strengths and limitations of each study, it is a serious error of omission to ignore such studies in reviewing the state of this science. (http://www.cognitivetrainingdata.org/the-controversy-does-brain-training-work/response-letter/)

Though it has no bearing on the scientific merits of working memory training specifically, there are strong indications that Lumosity in particular has overstated the efficacy of their ‘brain training’ products. In January 2016 the company was ordered to pay two million USD in order to settle charges of deceptive advertising, relating to claims of improved everyday performance, and protection against age-related cognitive decline (Federal Trade Commission). This is not to say that Lumosity training does not have the intended effects, but rather strongly suggests that the current state of the evidence is not robust enough to warrant the kinds of claims being made by the company.
Unfortunately, though not surprisingly, this polarization of the literature and scientific/professional opinion makes it difficult for interested readers to distill straightforward answers to the basic questions the literature originally sought to address, like ‘Does working memory training work?’, and, ‘If so, for whom does it work best, and under what conditions?’ Rather, to a large extent the literature is still grappling with the now century-old question of whether or not training in one task will generalize to ability gains on another. The broader question of whether or not people can be made smarter through any kind of training hangs in the balance. In short, resolving the issue of whether or not human intelligence is a fixed trait, and the degree to which cognitive processes relating to problem solving remain plastic over the lifespan hold massive practical importance because they inform foundational approaches to disciplines that not only support cognitive health, but also human flourishing: psychology, education, neurorehabilitation, and aging (Slagter, 2012). If humans can be trained to reason and plan more effectively, solve problems of greater complexity, think more abstractly, or comprehend complex ideas more fully by training WM, we urgently need to know about it. Conversely, it is similarly important to know if it is not possible to increase intelligence by training WM, so that our psychological, educational, and remedial efforts can be focused on strategies that augment our cognitive capacities by circumventing their limitations, rather than on ineffective exercises aimed at increasing them.

Given this background, the present set of studies seek to add clarity to the WM training literature by addressing a number of methodological shortcomings that have unfortunately become nearly ubiquitous amongst similar recent investigations (Button et al, 2013; Boot et al., 2013; Redick, 2015; Melby-Lervåg, Redick, & Hulme, 2016; Moreau et al., 2016). First, because the effects of WM training may be sensitive to the very process by which it is trained, our trial
includes an active control condition in addition to the WM training group and a no-contact control group. By having a separate group of participants complete similarly intensive cognitive training, we can be assured that any observed effects in the WM trained group are not due to placebo effects relating to interaction with experimenters or the training program itself. Second, each of these three groups in our trial contains just over the minimum suggested number of 20 participants (Simmons, Nelson, & Simonsohn, 2011). Including more than the standard 10-15 participants per group increases the power to find transfer effects should they truly exist, and reduces the likelihood that significant findings will be erroneous. Third, our trial measured a number of important intra-personal variables which have been suspected of moderating training effects in previous studies. Measuring salient variables such as motivation to complete training, expectations for improvement, personality traits (Ashton & Lee, 2009), grit (i.e. long-term determination; Duckworth & Quinn, 2009), need for cognition (Cacioppo, Petty, & Kao, 1984), as well as current cognitive activities (Eskes, et al., 2010) helps to reduce doubt as to whether one or several of these oft-overlooked variables accounts for important differences between groups. Finally, our trial is one of only a select few in the WM training literature to include a neuroimaging component. It is only the second such trial to do so with an active control condition, and to our knowledge, the first to scan participants for both the task they trained on, as well as the task to which their increased aptitude is purported to transfer.

Increasing the methodological rigor of the present trial through these means reduces the number of potential alternative explanations for the observed data, and uniquely situates us to be able to draw strong conclusions about the efficacy of WM training, as well as the neural mechanisms through which it is thought to occur. Adding one more experimental trial to the polarized literature on WM training cannot hope to settle the matter. However, our hope is to
increase confidence in important and straightforward answers to the ostensibly simple question of the current literature: ‘Does WM training work?’ In so doing, we hope to comment on the much older question about the malleability of human intelligence and cognition more generally.

These aims are accomplished through three separate manuscripts, each addressing a specific aspect of the questions outlined above. First, Clark, Lawlor-Savage, & Goghari (in preparation\textsuperscript{1}) compares the neural activation patterns associated with our functional MRI tasks of WM, fluid intelligence, and word decoding for all participants before completing any training. The manuscript reviews results of previous neuroimaging studies examining these cognitive domains through similar tasks, and compares the overlap in activation patterns. Second, Clark, Lawlor-Savage, & Goghari (in preparation\textsuperscript{2}) examines the cognitive outcomes of the WM trained group in relation to the active control condition, and the no-contact control group. Finally, Clark, Lawlor-Savage, & Goghari (in preparation\textsuperscript{3}) examines the pre- and post-training functional neuroimaging data for the WM training and active control groups.
Chapter Two: Comparing Brain Activations Associated with Working Memory and Fluid Intelligence

2.1 Background

Working memory (WM), or the ability to maintain and manipulate task-relevant information over short periods of time, is a core cognitive ability in humans. Though WM shares much in common with the concept of ‘short term memory’ (i.e. cognitive architecture, capacity limitation, and functional neuroanatomy), it is the prospective use of information in the service of some goal or objective which distinguishes the two, and has largely motivated the use of the descriptor “working” (Linden, 2007; Eriksson, Vogel, Lansner, Bergstrom, & Nyberg, 2015). In this sense, WM is highly utilized in a wide variety of challenging tasks both in the laboratory and in everyday life, and is in fact highly related to general intellectual and reasoning abilities, or ‘fluid intelligence’. In the theory of Cattell (1963), fluid intelligence (denoted \( G_f \)) is the ability to adapt one’s reasoning abilities to solve novel cognitive problems involving new information, and stands in contrast to ‘crystallized intelligence’ (denoted \( G_c \)) which draws heavily upon previously learned declarative information acquired from education or previous experience (Carroll, 1996; McGrew, 2009; Carpenter, Just, & Shell, 1990).

Fluid intelligence and WM are highly related psychological constructs, and are often described as ‘almost’ isomorphic. Studies investigating the specific strength of the relationship between WM and \( G_f \) have noted moderate correlations with coefficients in the .3 to .9 range (see Colom, Rebollo, Palacios, Juan-Espinosa, & Kyllonen, 2004; Burgess, Gray, Conway, & Braver, 2011), and Martínez et al. (2011) reported that WM and \( G_f \) could not in fact be distinguished at the latent variable level. Thus, although WM is defined much more narrowly than \( G_f \), research over the past several decades has identified it as a core psychological process responsible for
driving much of the observable variation in human cognitive abilities. Indeed, making recently experienced stimuli subsequently accessible for brief periods is an essential component of our ability to act outside the bounds of the immediate moment, and to coordinate complex goal-directed behaviours (Baddeley, 1992, Repovs & Baddeley, 2006). Given the relative necessity and ubiquity of WM processes in day-to-day cognitive functioning, previous research has indicated a surprisingly small capacity of the WM system – perhaps limited to only a few items (Chuderski, 2013). Luck and Vogel (1997) estimated the average capacity to be approximately four items, with most individuals demonstrating a capacity between two and six items (Cowan, 2001). Though small in absolute terms, these inter-individual differences in WM capacity appear to account for much of the variance observed in cognitively demanding tasks and situations, including reading comprehension, language abilities, mathematics, reasoning, problem solving, overall academic performance, and even ‘fluid intelligence’ more generally (Engle, Tuholski, Laughlin, & Conway, 1999; Eriksson et al., 2015).

A parallel approach to studying the psychometric association between WM and Gf is the investigation of the neural mechanisms associated with each, and their potential overlap in terms of observed patterns of activation in response to task-load (i.e. task difficulty/complexity). Shared components between tasks might represent capacity constraints in the domain of attention (see Halford, Cowan, & Andrews, 2007), or perhaps shared neural circuitry or cortical involvement between the two tasks (Buschkuehl, Hernandez-Garcia, Jaeggi, Bernard, & Jonides, 2014). Burgess and colleagues (2011) note that a better understanding of the core psychological and neural mechanisms involved in WM may help in developing better targets for intervention in training studies, and more accurate predictions about success or failure of such interventions. To this end, here we review literature examining typical activation patterns associated with two
commonly utilized tasks of WM and Gf: the dual n-back task, and Raven’s Standard Progressive Matrices (RSPM) respectively. In addition, we examine functional activation patterns associated with a comparison task that is theoretically unrelated to the domains of WM or Gf, in order to assess the specificity of observed activations for tasks in those domains. For this purpose, we chose a word/pseudoword decoding task (lexical decision task; LDT) that, contrary to fluid intelligence, draws heavily upon previously learned information (i.e. crystallized intelligence; Carroll, 1996; McGrew, 2009; Carpenter, Just, & Shell, 1990).

2.1.1 Neural Mechanisms Associated with the n-Back Task

The n-back task has been widely utilized in recent years to study WM both in the context of its neural bases (Owen et al., 2005; Rottschy et al., 2012), as well as how training with the task may serve to improve WM (see von Bastian & Oberauer, 2013 for a review). The n-back task requires participants to monitor consecutive presentation of visual and/or auditory stimuli, and respond via button-press when the current stimuli matches that presented ‘n’ trials ago, where n typically ranges from one to three. As n is increased, task difficulty rises sharply and places larger demands on a number of key processes involved in WM, including monitoring, updating, and manipulation of remembered information (Owen, McMillan, Laird, & Bullmore, 2005).

Owen and colleagues (2005) conducted a meta-analysis of 24 imaging studies utilizing the n-back task in healthy adult populations. Their analysis revealed five consistent areas of activation across task variants, falling exclusively within the frontal and parietal lobes: First, the bilateral dorsolateral prefrontal cortex (DLPFC) is thought to be involved in organizational control of WM, reducing overall cognitive load in WM tasks by selecting appropriate organizational chunks. Interestingly, neuropsychological data appears to support this claim, as
patients with damaged frontal lobes appear to be impaired on only some WM tasks (Owen, Morris, Sahakian, Polkey, & Robbins, 1996). Second, the bilateral mid-ventrolateral prefrontal cortex was implicated in n-back activity and is thought to be associated with explicit encoding and retrieval of information, as well as attentional processes. Third, the bilateral rostral prefrontal cortex is thought to combine or integrate multiple cognitive processes, specifically when the operation of a single cognitive process is insufficient to meet the particular demands of a task. Fourth, bilateral medial premotor cortex was implicated and is thought to be involved in maintenance of visuospatial attention during working memory tasks. Finally, bilateral medial posterior parietal cortex (including precuneus and inferior parietal lobule) was implicated in n-back activations and has previously been associated with mediating shifts in attention, retaining task-related temporal information, and preparing for a given task.

Similarly, Rottschy and colleagues (2012) undertook a broader neuroimaging meta-analytic approach to model the neural correlates of working memory by investigating several WM-related tasks (e.g. the Sternberg Task, the delayed matching to sample task, and the delayed simple matching task) as well as the n-back. They identified what they called a “core” WM network which included areas predominantly in the frontal and parietal regions across both hemispheres: dorsolateral prefrontal cortex, lateral prefrontal cortex, anterior insula, premotor cortex, pre-supplementary motor area, intraparietal sulcus, superior parietal lobule, and anterior parietal area. Interestingly, the authors also noted consistent activation bilaterally in ventral visual cortex, lobule VI of the cerebellum, and several subcortical areas including areas of the thalamus sharing connections with the prefrontal and temporal cortices, as well as predominantly left basal ganglia. Rottschy et al. (2012) note that their results are quite similar to the earlier findings of Owen et al. (2005), despite including a wider variety of WM tasks in their meta-
analysis and perhaps divergent naming practices of given stereotaxic spaces or neural structures (e.g. anterior insula implicated by Rottschy and colleagues versus frontal operculum implicated by Owen and colleagues). The consistency of these results across task and stimuli types strongly implicate the above areas as comprising a “core” WM network, perhaps forming the neural underpinnings of all WM cognitive processes.

Interestingly however, this core WM network closely resembles patterns of activation observed in a wide variety of tasks beyond WM, particularly when task demands increase. For example a similar, though right-dominant, network is associated with selective attention (see Shulman et al. 2002; Shulman et al., 2009), and largely similar networks studied outside the realm of intelligence have been termed the attention and working memory system (Cabeza & Nyberg, 2000), the cognitive control network (Cole & Schneider, 2007), and the task-positive network (Fox, Snyder, Vincent, Corbetta, Essen, & Raichle, 2005). The involvement of these same cortical areas in such a diversity of tasks suggest an even more fundamental role in cognition, and coordinating behaviour beyond WM. Indeed, the fronto-parietal network implicated in WM processes has also been discussed in terms of an executive control network, a dorsal attention network (Yeo et al. 2011), a core executive (discussed in Rottschy et al. 2012), and a multiple demand network (Duncan, 2010) which may mediate goal-directed behaviour by rapidly organizing mental focus and separation of successive task steps. In this sense, the cortical areas involved in successful navigation of WM tasks, or what Duncan (2010) calls ‘multiple-demand cortex’, may be essential to keeping the mind focused on the task-relevant information at hand without regard to its cognitive modality (e.g. perceptual, mnemonic, motor-related), thus playing a central role in virtually all non-routine cognitive functions. Recall this matches the traditional description of fluid intelligence quite well – i.e. the ability to adapt one’s reasoning
abilities to solve novel cognitive problems involving new information (Carroll, 1996; McGrew, 2009; Carpenter et al., 1990).

2.1.2 Neural Mechanisms Associated with Raven’s Progressive Matrices

Raven’s Progressive Matrices (RPM; Raven, 1975; Raven & Raven, 1994) scores are highly correlated with a wide range of other tests of intelligence, and is regarded as the most general single test of non-verbal (i.e. fluid) intelligence (Carpenter, Just, & Shell, 1990; Alderton & Larson, 1990). The task presents participants with a series of complex designs or grids of visual puzzle elements with a single piece missing, and requires that they choose the correct missing piece from several alternatives. Both the standard set of puzzles (Raven’s Standard Progressive Matrices; RSPM; Raven, 1976) and the more difficult set (Raven’s Advanced Progressive Matrices; RAPM; Raven, 1994) require participants to infer and integrate rules, manage goal hierarchies, as well as form abstractions on the basis of novel non-verbal information.

RPM-type tasks have not received as much attention as the n-back task in terms of neuroimaging studies, however early investigations have revealed activity in fronto-parietal regions. Prabhakaran and colleagues (1997) found increasing activation of the prefrontal cortex with progressively more complex, (and therefore more difficult) RPM problems. Consistent with this, Duncan and colleagues (2000) found increased DLPFC perfusion (assessed via positron emission tomography; PET) in response to performing a variety of Gf tasks, including RPM-type tasks. Two subsequent studies obtained similar results, and additionally found engagement of increasingly anterior areas of the prefrontal cortex with increased task difficulty (Christoff et al. 2001; Kroger, Sabb, Fales, Bookheimer, Cohen, & Holyoak, 2002). Selective activation of the DLPFC and posterior parietal cortex in these studies is thought to reflect the verbal working
memory and visuospatial processes utilized during the reasoning process. Similarly, activation of the rostrolateral prefrontal cortex is variously thought to support the more executive processes of working memory during reasoning, or the manipulation of self-generated information to support abstraction, or perhaps the integration of information from longer-term memory with the demands of the current task (reviewed in Krawczyk, 2012). More recent studies have further implicated additional brain areas in non-verbal reasoning in RPM tasks including the basal ganglia (Melrose, Poulin, & Stern, 2007), and the cerebellum (Kalbfleisch, Van Meter, & Zeffiro, 2007). Most recently, Shokri-Kojori, Motes, Rypmal & Krawczyk (2012) demonstrated increasing involvement of prefrontal cortex, as well as occipital and parietal cortices as task complexity increased from one to three relations in an RPM-type visuospatial reasoning task.

Two studies have investigated neural activation differences for RPM in high- and low-ability healthy adult samples. Perfetti and colleagues (2009) found that high-ability participants responded to increasing task complexity with commensurate increases in activation in frontal and parietal areas, whereas low-ability participants showed decreased activation in the same areas. Further, low-ability participants demonstrated greater activity overall in response to less difficult problems, particularly in medial and lateral frontal areas. The authors suggest that these patterns reflect the discrepancy in mental effort required between the groups, and indicate considerable differences in problem solving strategies, executive control, and/or modulation of attention as task demands increase. Lee and colleagues (2006) found similar results in comparing high- and low-ability participants, though noted relatively stronger differences in posterior parietal cortex.

2.1.3 Neural Mechanisms Associated with the Lexical Decision Task

By including a task clearly outside the domain of WM or Gf (i.e. semantic knowledge, or vocabulary, or ‘crystalized intelligence’), we can gain a better understanding of the degree to
which tasks of WM and/or Gf exclusively activate whichever regions they are found to. The lexical decision task presents participants with four letters, and requires that they decide as quickly as possible whether these letters constitute a ‘real’ word in English, or merely a pronounceable pseudoword. Though not the specific cognitive domain or neural activation pattern of interest, this task was included here as a comparison to speak to the specificity of potential overlapping neural activations for the WM and Gf tasks. Unfortunately however, little is known about neural activation patterns associated with the lexical decision task when contrasting different levels of task difficulty in blocked designs. The vast majority of fMRI-based neuroimaging studies on the lexical decision task have utilized event related designs, and investigated real word versus pseudoword contrasts (Taylor, Rastle, & Davis, 2013). Haut, Lim and MacDonald (2010) did contrast easier and harder blocks of this task, though only reported on null findings for cortical areas of change following cognitive intervention in a sample of participants with schizophrenia. In this sense, our findings for patterns of functional neural activation in response to task load for the lexical decision task are novel, though will still serve the purpose of speaking to the specificity of findings for the two central tasks of interest (i.e. dual n-back, and RSPM).

2.1.4 The Neuroanatomy of Intelligence

The proposed roles for frontal and parietal brain areas associated with the n-back task and RPM are highly consistent with Jung & Haier’s (2007) Parieto-Frontal Integration Theory (P-FIT) of intelligence, which attempts to localize general cognitive ability within the brain via evidence from neuroimaging studies. According to P-FIT, auditory and visual information is collected by the temporal and occipital lobes, while the majority of processing on this information (i.e. that which separates those of high and low intellectual ability) occurs in a
complex interplay between frontal and parietal cortical areas. Specifically, sensory information from the temporal and occipital lobes is fed forward to bilateral parietal cortex leading to the emergence of abstraction, elaboration, and symbolism. From there, task relevant information is fed forward to frontal regions which ‘hypothesis test’ various solutions to the given problem, and work to inhibit conflicting solutions. Importantly, the entire circuit is dependent on error-free transmission of information between cortical areas via white matter tracts, notably the arcuate fasciculus. The P-FIT model is consistent with lesion mapping studies evincing significant associations between damage to ‘remarkably circumscribed’ networks of frontal/parietal cortices and decreases in intellectual abilities (Glascher et al., 2010; Barbey, Colom, Soloman, Krueger, Forbes, & Grafman, 2012).

Recent support for the P-FIT model comes from a voxel-based quantitative meta-analysis of structural and functional imaging studies of cognitive ability. Basten and colleagues (2015) integrated neuroimaging data from 28 separate neuroimaging studies, and their results were highly consistent with the P-FIT model. However, results of their analysis suggested somewhat more emphasis on the importance of the frontal and parietal cortices versus sensory cortices (i.e. temporal/occipital) compared to the original model. Overall, this body of work indicates that there are indeed regions in the healthy adult brain that vary in structure and function from person to person based on intellectual ability. This strongly suggests a neurological substrate for general intelligence.

The strong behavioural and psychometric associations between general intelligence, WM, and Gf suggest similar functional networks for these constructs. However, relatively few studies have worked to directly and accurately characterize the similarities and differences between the networks responsible for WM and Gf which support critical cognitive skills such as adaptive
behaviour and novel problem solving. In line with the expectation that WM and Gf should also rely to a large extent on the fronto-parietal network, Gray and colleagues (2003) found that participants’ DLPFC activation during a difficult (i.e. high-interference) 3-back task was the best predictor of Gf as measured by RPM. Follow-up analyses further revealed that lateral prefrontal and parietal regions mediated the relation between Gf and performance. Most recently, lesion mapping studies have indicated that impairments in WM and Gf were associated with a circumscribed right-lateralized fronto-parietal cortical network, as well as the connecting white matter tracts, superior longitudinal/arcuate fasciculus (Barbey, Colom, Paul, & Grafman, 2014). Interestingly, earlier lesion mapping work by these authors indicated a mirrored left-lateralized fronto-parietal network to be associated with deficits in general intelligence (Barbey et al., 2012).

Given this background, the central goal of the present investigation is to further clarify the nature of the neural networks that support WM versus Gf processes in human cognition. To this end, we studied a relatively large sample of 63 healthy adult participants in an fMRI design. Importantly, by utilizing in-scanner tasks that have become ubiquitous in the working memory training literature (i.e. dual n-back task, and RPM), our results are uniquely positioned to comment on the potential overlap in functional activation for WM training (i.e. dual n-back) tasks and potential transfer to tasks of Gf (i.e. RPM).

2.2 Method

2.2.1 Participants and Recruitment

A total of 63 healthy adult participants between the ages of 18 and 40 were recruited from the community through advertisements and physical postings distributed throughout the community. Exclusion criteria included: 1) left-handedness; 2) history of traumatic brain injury or other neurological condition causing sensory or motor impairment; 3) presence of self-
reported mental illness; 4) less than normal or corrected-to-normal visual acuity; and 5) MRI contra-indications. Recruitment was shared with that of larger investigation of online working memory training, thereby necessitating the additional exclusion criteria of: 6) insufficient access to a computer and high-speed internet; and 7) recent or previous use of the n-back training task or other online cognitive training paradigms. All potential participants were directed to online surveys of demographic, behavioural, and health information to confirm eligibility. Participants were compensated $20 per cognitive testing session, and $20 per MRI session. Written consent was obtained from all participants, and ethics approval was obtained from the University of Calgary’s Conjoint Health Research Ethics Board (CHREB).

2.2.2 Procedure and Materials

Following initial recruitment and eligibility screening, participants underwent MRI scanning. MRI sessions consisted of structural scans for anatomical registration, and functional scans of the dual n-back task, Raven’s Standard Progressive Matrices, as well as a lexical decision task. Participants were introduced to these tasks before the MRI session, and were asked to practice with them until they indicated full comprehension of task instructions and requirements. These practice trials included immediate feedback for responses, whereas the trials in the scanner did not. In the scanner, participants viewed stimuli on a screen suspended behind their heads in the magnet bore via a mirror attached to the head coil. Participants were additionally fitted with headphones for presentation of audio stimuli, and utilized a scanner-compatible two-button response box placed in their right hand in order to indicate their responses to task trials with their index and/or middle fingers. All stimuli were presented electronically using E-Prime software (Psychology Software Tools, 2012). MRI sessions typically ranged from
60 to 75 minutes in duration. The three tasks described below were completed in a random order for each participant.

2.2.2.1 Dual n-Back Functional Imaging Task

Jaeggi and colleagues’ (2008) description of the dual n-back training task was followed as closely as possible, though modified for use in the scanner. For each trial in a block-design, participants viewed a blue square presented at one of eight locations around the periphery of a 3x3 grid with a fixation cross at centre, and simultaneously listened to an auditorily presented letter (D, F, H, J, M, Q, R, or X). Visual stimuli were presented for 500 milliseconds (ms), followed by 2500 ms of fixation which also served as the window within which participants could indicate their responses for that trial. Participants indicated a visual match to n trials previous via button-press with their index finger, auditory match to n trials previous via button-press with their middle finger, and combined visual and auditory match via button-press with both fingers. No response was required on trials not containing a match. Participants underwent four blocks each of 1-back, 2-back, and 3-back conditions, where each block included 20+n trials. Visual stimuli location, and letter presentation were randomized within each block, though each block was structured to contain four visual matches, four auditory matches, and two combined visual plus auditory matches. Each block was preceded by 5000 ms of instructions, and separated from subsequent blocks by 15 seconds of fixation in order to allow the hemodynamic response to return to baseline. Block order was fixed for all participants: 1-back, 2-back, 3-back repeated four times, split into two separate scan runs lasting 8:51 each, for a total task time of 17 minutes, 42 seconds.
2.2.2.2 Raven’s Progressive Matrices Functional Imaging Task

Raven’s standard set of matrix problems (Raven, 1976) were modified for use in the scanner such that participants could respond using the same scanner-compatible two-button response box utilized for the n-back task. This was achieved within a block-design by first presenting participants with each matrix problem for 5000 ms, after which one of the multiple answer options was highlighted with a superimposed red box. Participants were then required to indicate whether this highlighted answer was correct via button press with their index finger, or incorrect with their middle finger. The matrix problem and highlighted answer option remained on the screen for 8000 ms, and trials were separated by 1000 ms of fixation cross at the centre of the screen. The highlighted answer was the correct answer for exactly half of the trials, and the increased difficulty of ‘good wrong answers’ (i.e. those that closely resemble the actual correct answer) was controlled by ensuring an equal number in each block. The total set of 60 matrix problems that comprise Raven’s Standard Progressive Matrices was split into two sets (for pre-versus post-training comparisons as part of the larger working memory training study), and further divided into three levels of difficulty: Easy, Medium, and Hard – yielding six sets of 10 problems each. Participants completed one block of each difficulty level in a randomized order, in a single scanner run. Trials within blocks were also randomized. Task instructions were displayed for 5000 ms before the first block, and each block was preceded by 15 seconds of fixation in order to allow the hemodynamic response to return to baseline. Total task time was eight minutes.

2.2.2.3Lexical Decision Functional Imaging Task

For this block-design task, participants viewed sequences of four letters, and were asked to indicate whether they composed a real word in English via button-press; index finger for ‘yes’,
middle finger for ‘no’. Four-letter combinations were generated ahead of time using The English Lexicon Project database (Balota et al., 2007), which classifies words and non-words along several empirically derived dimensions, including accuracy of lexical decision (i.e. word versus non-word). For example, ‘blue’ and ‘jume’ have perfect accuracy ratings in the database, indicating that they are relatively easy to distinguish as a word and non-word respectively. In contrast, ‘faze’ and ‘thew’ have considerably lower accuracy ratings reflecting their nature as harder series’ of letters to lexically classify. Participants completed three blocks each of Easy and Hard four-letter combinations, in a single scanner run. Blocks consisted of 30 trials each in which half were real words and half were non-words. Block order was randomized, as were items within each block. Task instructions were displayed for 5000 ms before the first block, and each block was preceded by 15 seconds of fixation in order to allow the hemodynamic response to return to baseline. Total task time was 10 minutes, 45 seconds.

2.2.3 MRI Data Acquisition

Whole-brain images were collected for all participants on a 3T General Electric Discovery MR750 system using an 8-channel head coil at the Seaman Family Magnetic Resonance Research Centre at the University of Calgary. Functional echo planar imaging (EPI) data were acquired in an interleaved bottom-up slice order with 40, 3.4 mm thick slices, echo time (TE) = 30 ms, repetition time (TR) = 2500 ms, flip angle = 77°, field of view (FOV) = 22, and matrix = 64 × 64. A total of 424 volumes were collected for the n-back task, 192 volumes for Raven’s Progressive Matrices, and 258 for the lexical decision task. T1-weighted 3D magnetization-prepared rapid gradient-echo (MP-RAGE) anatomical scans were also acquired for each participant to register the functional data (256, 1 mm slices, TE = 3.1 ms, TR = 7.4 ms, inversion time (TI) = 650 ms, FOV= 25.6, matrix = 256 × 256). Additionally, T2-weighted high
resolution scans were collected as expanded functional images to aid in registering the functional images to standard space (40, 3.4 mm thick slices, TE = 120 ms, TR = 7500 ms, FOV = 22, and matrix = 256 × 256).

2.2.4 Data Analysis

Functional MRI data was analysed with fMRI Expert Analysis Tool (FEAT) and other utilities from FMRIB Software Library (FSL; www.fmrib.ox.ac/fsl) version 5.0.9 for Linux. Preprocessing of functional and anatomical images included reorientation to standard orientation, removal of non-brain tissue, slice-timing correction (Smith, 2002), spatial smoothing (using a 7 mm FWHM kernel), grand-mean intensity normalization, and high-pass temporal filtering. Time-series statistical analyses were carried out using FMRIB’s improved linear model (FILM) with local autocorrelation correction (Woolrich 2001). Careful inspection of plots representing head movement revealed less than 2 mm of translation for all participants, so motion correction parameters were not included in the preprocessing stage. Functional images were registered first to T2-weighted high resolution images, then to participants’ structural images, and then to standard the MNI avg152 T1-weighted template with non-linear transformations with seven, seven and 12 degrees of freedom respectively using FSL’s FMRIB Linear Image Registration Tool (FLIRT; Jenkinson & Smith, 2001; Jenkinson, Bannister, Brady, & Smith, 2002). Manual inspections were performed after each preprocessing step for each scan, and reanalyses were performed where necessary. Mean group activations for all harder > easier task condition contrasts were modeled for each task contrast using a Z threshold of 2.3 and a cluster threshold of 0.05 (Worsley, 2001). Two task scans were lost due to problems in data acquisition. Therefore, the group mean analyses were based on 62 participants for the dual n-back task, 63 participants for RSPM, and 62 participants for the lexical decision task.
Participant behavioural responses to task stimuli in the scanner were collected via E-Prime software, and exported to SPSS for analysis. Specifically, one-way within-subject ANOVAs were utilized to assess for significant differences in performance and reaction times across task difficulty levels for the n-back task and Raven’s Standard Progressive Matrices (e.g. 1-back, 2-back, 3-back; Easy, Medium, Hard matrix problems). A dependent samples t-test was used to assess for analogous differences between the two difficulty levels of the Lexical Decision Task (Easy versus Hard).

2.3 Results

2.3.1 Participant Demographics and Cognitive Characteristics

The 63 participant sample was composed of 28 males and 35 females, with an average age of 30.91 (SD = 6.01), and an estimated average educational attainment of 15.44 (SD = 1.90) years. Cognitive testing revealed intelligence quotients (IQ) ranging from 69 to 143 with an average of 109.95 (SD = 14.93) based on scores from eight split-half WAIS-IV subtests.

2.3.2 In-Scanner Behavioural Data

As expected, increasing task difficulty resulted in fewer correct responses from participants across all tasks: n-back, \(F(2, 124) = 231.76, p < .001\); RSPM, \(F(2, 124) = 38.93, p < .001\); Lexical Decision Task, \(t(62) = 7.54, p < .001\). Follow-up pairwise analyses using the Bonferroni correction revealed that correct responses for each of the three difficulty levels of the n-back task were in fact significantly different from each other: 1-back, 143.83 ± 1.96 total correct responses (of a possible 160); 2-back, 133.95 ± 2.03; 3-back, 124.65 ± 1.42. However, follow-up pairwise analyses for the RSPM task revealed significant differences in participant performance only between the easiest level of difficulty and the harder two. On average, participants answered 9.21 ± 0.29 (of a possible 10) easy matrix problems correctly, versus only
7.68 ± 0.28 medium questions ($p < .001$ compared to easy), and $7.91 ± 0.29$ hard questions ($p < .001$ compared to easy). Performance on medium and hard matrix problems was statistically indistinguishable ($p = .60$). Increasing task difficulty also increased reaction times for both the RSPM task [$F(2, 124) = 63.55$, $p < .001$], and the Lexical Decision Task $t(63) = -15.45$, $p < .001$. Interestingly, follow-up pairwise analyses of reaction times, again using the Bonferroni correction for the RSPM task revealed statistically different reaction times for each of the three levels of difficulty ($p < .001$ for each pairwise comparison). Reaction time was not considered for the n-back task, as on exactly half of all trials, the correct response was to not press a button. See Figure 1 for a graphical representation of task performance and reaction time.
Figure 1. Performance and reaction time data for in-scanner tasks: (A) dual n-back task, (B) Raven’s Standard Progressive Matrices, and (C) a Lexical Decision Task. Error bars represent 95% confidence intervals.
2.3.3 Whole Brain fMRI Analyses

2.3.3.1 Dual n-Back Task, by Level of Difficulty

2.3.3.1.1 3-back > 1-back Contrast

Mean group activation was observed when contrasting the hardest and easiest conditions of the task – specifically in the bilateral middle and superior frontal gyri, paracingulate gyri, frontal poles, frontal orbital cortex, insular cortex, precuneus, superior parietal lobule, supramarginal gyrus, angular gyrus, the left caudate, and areas of the right cerebellum. This suggests that the WM network discussed above is indeed activated by the present task, even at the hardest level of difficulty with relatively poor performance. See Table 1 and Figure 2.

Table 1.

Significant whole-brain activations by task and contrast

<table>
<thead>
<tr>
<th>Task and Contrast</th>
<th>Regions within each cluster</th>
<th>Voxels</th>
<th>Z_max</th>
<th>MNI coordinates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dual n-Back Task</td>
<td>3-back &gt; 2-back</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>R/L middle frontal gyrus, R/L paracingulate gyrus, R/L superior frontal gyrus, R/L frontal pole, R/L frontal orbital cortex, R/L insular cortex, L caudate</td>
<td>26438</td>
<td>7.67</td>
<td>-6 28 38</td>
</tr>
<tr>
<td></td>
<td>R/L precuneus, R/L superior parietal lobule, R/L supramarginal gyrus, R/L angular gyrus</td>
<td>14772</td>
<td>7.02</td>
<td>6 -68 54</td>
</tr>
<tr>
<td></td>
<td>R cerebellum crus I, R/L cerebellum crus II</td>
<td>1503</td>
<td>5.04</td>
<td>34 -62 -34</td>
</tr>
<tr>
<td>2-back &gt; 1-back</td>
<td>R/L inferior temporal gyrus, R/L middle frontal gyrus, R/L precentral gyrus, R/L paracingulate gyrus, R/L superior frontal gyrus, R/L frontal pole, R/L frontal orbital cortex, R/L insular cortex, R/L precuneus, R/L superior parietal lobule, R/L middle temporal gyrus, R/L supraclecarine cortex, R/L intracalcarine cortex, R/L lingual cortex, R/L cuneal cortex, occipital pole, R/L occipital fusiform gyrus, R/L caudate, R/L thalamus, R/L cerebellum</td>
<td>93614</td>
<td>8.85</td>
<td>-50 18 26</td>
</tr>
</tbody>
</table>

Raven’s Std. Matrices
<table>
<thead>
<tr>
<th>Condition</th>
<th>Area</th>
<th>Z-Score</th>
<th>X</th>
<th>Y</th>
<th>Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hard &gt; Medium</td>
<td>R inferior frontal gyrus, R superior frontal gyrus</td>
<td>2.023</td>
<td>4.05</td>
<td>44</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>L inferior temporal gyrus</td>
<td>1.653</td>
<td>4.64</td>
<td>-52</td>
<td>-60</td>
</tr>
<tr>
<td>Hard &gt; Easy</td>
<td>L lateral occipital cortex, L occipital pole</td>
<td>2.225</td>
<td>4.14</td>
<td>-34</td>
<td>-88</td>
</tr>
<tr>
<td></td>
<td>L superior parietal lobule</td>
<td>1.289</td>
<td>4.98</td>
<td>-26</td>
<td>-68</td>
</tr>
<tr>
<td></td>
<td>R occipital pole</td>
<td>1.169</td>
<td>4.84</td>
<td>38</td>
<td>-92</td>
</tr>
<tr>
<td>Medium &gt; Easy</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Lexical Decision Task</td>
<td>R/L thalamus, R/L putamen, R/L pallidum,</td>
<td>2.291</td>
<td>5.87</td>
<td>-8</td>
<td>-10</td>
</tr>
<tr>
<td>Hard &gt; Easy</td>
<td>R/L caudate, R inferior frontal gyrus, R middle frontal gyrus, R frontal pole, R insular cortex, R frontal orbital cortex, R/L lateral occipital cortex</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>R/L cingulate gyrus, R/L paracingulate gyrus, R/L superior frontal gyrus, L inferior frontal gyrus, L middle frontal gyrus, L frontal pole, L insular cortex, L frontal orbital cortex, L temporal pole</td>
<td>1.160</td>
<td>7.16</td>
<td>-50</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>L lateral occipital cortex, L superior parietal lobule, L supramarginal gyrus, L angular gyrus</td>
<td>1.289</td>
<td>4.86</td>
<td>-22</td>
<td>-68</td>
</tr>
<tr>
<td></td>
<td>R lateral occipital cortex, R superior parietal lobule, R supramarginal gyrus, R angular gyrus</td>
<td>1.139</td>
<td>4.78</td>
<td>34</td>
<td>-60</td>
</tr>
</tbody>
</table>
Figure 2. Significant whole brain activation for contrast 3-back > 1-back of the Dual n-Back Task.

2.3.3.1.2 3-back > 2-back Contrast

However, when comparing the more difficult 3-back condition to the 2-back condition, this WM-related network was not observed. Mean group activation was not observed to increase
commensurate with the increase in task difficulty, revealing no significant areas of activation for this contrast.

2.3.3.1.3 2-back > 1-back Contrast

Mean group activation was observed to increase in a wide variety of areas as a function of WM load (i.e. task difficulty) in comparing the 2-back and 1-back task conditions. These areas were consistent with the ‘core WM network’ discussed above, including large bilateral frontal and parietal regions, bilateral basal ganglia and thalamus, as well as more circumscribed regions of the temporal and occipital cortices, and cerebellum. See Table 1 and Figure 3.
Figure 3. Significant whole brain activation for contrast 2-back > 1-back of the Dual n-Back Task.
2.3.3.2 Raven’s Progressive Matrices Task, by Level of Difficulty

2.3.3.2.1 Hard > Medium Contrast

Mean group activation was observed to increase as a function of task difficulty in comparing the hardest and intermediate matrix problem sets. Specifically, in right frontal and left inferior temporal cortical areas. See Table 1 and Figure 4.

*Figure 4. Significant whole brain activation for contrast Hard > Medium of the Raven’s Standard Progressive Matrices Task.*
2.3.3.2.2 Hard > Easy Contrast

Similar to the comparison of the hardest and intermediate matrix problems, comparison of the hardest versus easiest problems indicated increased activation in left lateral occipital cortex, left superior parietal lobule, and bilateral occipital poles. See Table 1 and Figure 5 for more detailed representations of these activations.

*Figure 5.* Significant whole brain activation for contrast Hard > Easy of the Raven’s Standard Progressive Matrices Task.
2.3.3.2.3 Medium > Easy Contrast

Easy and Medium matrix problems were indistinguishable on the basis of their group level activations, suggesting that the two groupings may not have engaged significantly different cognitive processes.

2.3.3.3 Lexical Decision Task, by Level of Difficulty

2.3.3.3.1 Hard > Easy Contrast

Consistent with the n-back and RSPM tasks, mean group activations were observed to increase as a function of task difficulty for the lexical decision task. Activation increases occurred in a wide variety of regions including bilateral medial and lateral frontal lobes, insula, anterior cingulate, lateral occipital cortex, basal ganglia, cerebellum, as well as left superior parietal cortex. See Table 1 and Figure 6 for more detailed representations of these activations.
Figure 6. Significant whole brain activation for contrast Hard > Easy of the Lexical Decision Task.

2.4 Discussion

Functional MRI of the tasks most typically utilized in the literature for WM training (i.e. dual n-back task), and as a gauge of transfer to Gf (i.e. RPM-type tasks) revealed considerable overlap in neural activation in response to increasing task demands. Similar areas of the frontal
and parietal cortices were also activated by increasing task demands in our comparison lexical decision task which is not thought to implicate WM or Gf cognitive processes at all, but rather relies on previously learned declarative information, or crystallized intelligence. To our knowledge, the present study is the first to investigate neural activations associated with the dual n-back and RSPM tasks in a single sample. Specifically, and consistent with previous literature, increasing \( n \) in the dual n-back task from one to two was associated with widespread activation increases in frontal and parietal cortices bilaterally, as well as basal ganglia, thalamus and cerebellum in our relatively large sample of healthy adults. Each of the frontal and parietal areas identified as part of the ‘core WM network’ by comprehensive meta-analyses of n-back tasks (Owen et al., 2005; Rottschy et al., 2012) were included in our very large activation cluster for this contrast, including: dorsolateral prefrontal cortex, lateral prefrontal cortex, anterior insula, premotor cortex, pre-supplementary motor area, intraparietal sulcus, superior parietal lobule, and anterior parietal area. Furthermore, present findings are consistent with the small number of studies which have investigated functional activation in response to simultaneous visual-spatial and/or auditory-verbal n-back task (i.e. dual n-back; Buschkuehl et al., 2014; Jaeggi, Buschkuehl, Etienne, Ozdoba, Perrig, & Nirkko, 2007; Yoo, Paralkar, & Panych, 2004; Thompson, Waskom, & Gabrieli, 2016). Interestingly however, increasing \( n \) from two to three was not associated with additional areas of activation despite the observed difference in performance accuracy between the two conditions. This is perhaps not surprising because failing to accurately complete the 3-back condition due to WM capacity constraints is, in theory, quite similar to completing an easier version of the task. That is, above one’s individual WM capacity threshold, all more difficult versions of the task may in fact evoke activation similar to that observed at that threshold. This is an empirical question that was not specifically tested in the
current experimental design. Other studies have investigated n-back task conditions greater than \( n = 3 \) however. For example, Buschkuehl et al. (2014) included a 4-back auditory condition in their investigation, and Schweizer et al. (2013) included a 5-back condition in a modified emotionally-laden version of the n-back task. It should be noted here that the lack of significant activation observed for the group average contrast between the 3-back and 2-back conditions may obscure variability in ability and activation patterns for different individuals within the group. For example, Jaeggi et al. (2007) found that as n-back task demands grew, lower-ability participants demonstrated load dependant *increases* in activation, whereas higher-ability participants evinced less activation while continuing to perform the task accurately.

From a cognitive perspective, failure to complete higher difficulty conditions of the n-back task may be accounted for by a failure of any of the involved component processes that the task is thought to require. For example, 3-back task requirements may 1) exceed the ability of participants to store or update individual spatial location and auditory letter items or sequences in memory (i.e. primary memory); 2) exceed the ability of participants to effectively search for target items that may be stored in memory (i.e. secondary memory); or 3) may exceed the ability of participants to effectively attend only to relevant task stimuli (i.e. attentional control; Unsworth, 2014). Based on estimates of average WM capacity between two and six items (Luck & Vogel, 1997; Cowan, 2001) it is plausible that any or all of the above constraints limit participant performance accuracy (and consistent functional activation) at the 3-back level of difficulty. Less difficult levels of the n-back task appear not to have exceeded participants’ abilities in these cognitive subdomains of WM.

The present findings for n-back task activations are also consistent with several of the finer differences between WM task characteristics noted by Rottschy and colleagues (2012) in
their meta-analytic review. For example current results are expectedly more similar to the bilateral WM load-dependent effects (i.e. more versus less difficult WM tasks) observed in ventral premotor cortex, lateral prefrontal cortex, inferior frontal gyri, supplementary motor area, and middle cingulate cortex, versus the primarily left-hemispheric set-related network (i.e. WM tasks vs. control tasks) observed in left rostral prefrontal cortex, superior parietal lobule/intraparietal sulcus, postcentral sulcus, dorsal premotor cortex, and posterior superior frontal gyrus. Rottschy et al. (2012) also split the corpus of WM tasks in their meta-analysis with respect to the type of recall required: verification (i.e. indicating whether a probe is the same as a memorized item), matching (i.e. indicating which of several probes is the same as a memorized item), or reproduction (i.e. replication of a previously memorized item). Verification experiments were found to produce stronger activations in the left intraparietal sulcus and superior parietal lobule compared to the other two, which showed greater activations in right inferior frontal gyrus (matching), and left posterior superior frontal gyrus (reproduction). Present results partially corroborate this distinction by evincing a relative lack of activation in the inferior frontal gyri, and bilateral activation of the superior parietal lobule for the 2-back > 1-back contrast. Greater convergence with previous literature was observed in comparing present results to the distinctions between task type, in Rottschy and colleagues meta-analysis. Specifically, n-back tasks were observed to activate bilateral intraparietal sulcus, anterior insula, posterior superior frontal gyrus, and the lateral prefrontal cortex, while Sternberg WM tasks tended to activate left basal ganglia, and left inferior frontal gyrus. Rottschy and colleagues (2012) also examined verbal versus nonverbal implementations of WM tasks, as well as object identity versus object location implementations of WM tasks, however, current results are more difficult to meaningfully compare to these distinctions as the present dual n-back task utilized
both verbal (i.e. auditorily presented letters), and nonverbal (i.e. blue squares) stimuli, and also requires participants to recall object identity (i.e. specific letters), and location (i.e. blue square location). In order to ascertain which of the above areas were consistently activated in WM tasks regardless of task specific features or implementations, Rottschy et al. performed a conjunction analysis across all these separate distinctions. Their analysis revealed a small set of loci that they implicate as the highly stable WM “core” network, including bilateral dorsal area 44 and premotor cortex, anterior insula, pre-supplementary motor area, intraparietal sulcus, left inferior parietal cortex, and right lateral prefrontal cortex.

Also consistent with previous literature, analysis of the RSPM task revealed much smaller areas of increased frontal and parietal activation in response to increasing task difficulty. However as discussed above, the Medium and Hard levels of the task were behaviourally differentiated only by reaction time, not task performance. Regardless, the increase in difficulty from Medium to Hard revealed increased activation of the right lateral inferior and superior frontal gyri, as well as the left lateral inferior temporal gyrus. Contrasting the Hard and Easy conditions additionally revealed increased activation of the occipital poles bilaterally, and the left superior parietal lobule. The overall smaller area of activation observed for the RSPM task compared to the n-back task may have arisen from the less challenging nature of the task altogether, or perhaps from the more gradually graded levels of difficulty compared to the present implementation of the n-back task. For example, the difference between 1-back and 2-back conditions of the n-back task is likely much larger than that of the difference between easy and medium problem sets of RSPM, particularly for our well-educated, above average intelligence sample. Consequently, comparing more similar difficulty levels of the RSPM task may have resulted in ‘cancelling out’ areas of activation that would not have occurred if the
problem set were split into two rather than three levels of difficulty (see below for further discussion).

Regardless of this potential difference, of particular interest here is the degree to which observed activation patterns for these tasks of WM and Gf converge, and where they do not. As can be seen in Figure 7, panel A, the areas of activation observed for the RSPM task are largely subsumed by the much larger activation changes observed for the dual n-back task, with the exception of left lateral occipital cortex, and the left occipital poles.

Figure 7. Harder > Easier difficulty contrasts for all tasks overlaid. Panel A shows the overlap of the dual n-back task (blue) and Raven’s Standard Progressive Matrices (green); and panel B shows the overlap of the dual n-back task (blue) with the lexical decision task. Images displayed in radiological convention.

Interestingly however, this pattern matches that which might be predicted by the task differentiation analyses performed by Rottschy and colleagues (2012). For example, in
comparing different types of WM task implementation, they found that matching of memorized items (versus verification or reproduction) in a WM task relied to a greater extent on right inferior frontal gyrus – a precise area implicated in the Hard > Medium task difficulty contrast for the present task, but not implicated in the larger 2-back > 1-back contrast, nor the 3-back > 2-back contrast for the n-back task. In this context, the current implementation of the RSPM task might be thought of as a primarily visuospatial, nonverbal, object location centered WM task (versus the traditional conceptualization of it as a test of Gf; Carpenter, Just, & Shell, 1990; Alderton & Larson, 1990) that requires matching of memorized stimuli in order to produce the correct answer on a given trial. This is consistent with Martínez and colleagues’ (2011) latent-variable approach to behavioural cognitive testing data indicating not only that short-term storage, WM, and WM updating are ‘hardly distinguishable’, but also that Gf itself is almost perfectly correlated with these cognitive abilities. This is perhaps not surprising upon closer inspection of RSPM task demands. In visually searching through the problem set for any given item of the RSPM task, participants must populate a list of relations amongst all permutations of adjacent matrix items, temporarily hold those relations in memory, and subsequently search the multiple potential answers looking for a match for these relations to rule out incorrect answers and find the correct one. In this sense, patterns of neural activation for the dual n-back and RSPM tasks show good convergence with their high degree of overlap from a cognitive (Burgess et al., 2011) and psychometric perspective (Chuderski, 2013) despite typically being thought of as tasks representative of distinct cognitive domains (i.e. WM and Gf).

Another way to conceptualize this overlap in neural activation for these cognitively distinct, though related, tasks is to appreciate their shared reliance on key areas of the frontal and parietal cortices not dissimilar from what has been described as the core WM network (Rottschy
et al., 2012; Owen et al., 2005), or any of the other monikers it has been given to highlight its importance in cognition more broadly: *attention and working memory system* (Cabeza & Nyberg, 2000), *the cognitive control network* (Cole & Schneider, 2007), *the task-positive network* (Fox, Snyder, Vincent, Corbetta, Essen, & Raichle, 2005), the *executive control network*, or the *dorsal attention network* (Yeo et al. 2011), *multiple demand network* (Duncan, 2010), or more simply, those areas outlined by Jung & Haier’s (2007) P-FIT model of intelligence. Similar regions have also been implicated in executive functioning related measures of task switching (Wager, Jonides, & Reading, 2004), inhibition (Nee, Wager, & Jonides, 2007), as well as lower-level cognitive functions including motor tasks involving planning (Bortoletto and Cunnington, 2010), movement integration (Wolynski, Schott, Kanowski, & Hoffmann, 2009), and even orientation (Marangon, Jacobs, & Frey, 2011). Indeed, the centrality of these select areas of cortex to such a vast range of cognitive activity support the notion that this “executive committee” network (Baddeley, 1996) may play an even more fundamental role in all human cognitive processes outside the circumscribed domains of WM and Gf. This nearly-ubiquitously engaged fronto-parietal network has even been described as essential in keeping the mind focused on task-relevant and goal-directed information regardless of task-type (e.g. perceptual, mnemonic, or motor-related), and even as mediating the control of goal directed behaviour more generally (Rottschy, et al., 2012). Rottschy and colleagues end their review on this important note regarding the centrality of the WM core network in all cognition: “We would hence conclude that the robustly engaged fronto-parietal core network, as demonstrated here for working memory, may sustain extremely basal processes or computations that are required for virtually all cognitive function, but whose exact nature remains to be further elucidated” (pp. 18). Thus, while from a behavioural perspective the dual n-back and RSPM tasks may appear completely
different and representative of distinct domains of cognitive functioning, neuroimaging evidence presented here and elsewhere suggests that the brain may treat RSPM-type problems more parsimoniously as simply another type of WM/attention/executive control task.

Further evidence of the centrality of this core WM network in widespread areas of cognition is provided by our lexical decision task. While classifying letter strings into categories of words or non-words would not typically be considered a task of WM or certainly Gf, our analyses revealed activation in areas quite similar to those of the dual n-back task and RSPM when task demands increased, including: bilateral medial frontal cortex, insula, anterior cingulate, basal ganglia, as well as left superior parietal areas (see Figure 7, panel B). Despite this apparent overlap in neural activation, decades of cognitive psychology would suggest that these tasks are in fact representative of very discrepant domains of psychological functioning (i.e. crystallized intelligence versus fluid intelligence; Gottfredson & Saklofsky, 2009). Rather than to suggest that lexical decision making involves similar psychological processes to dual n-back or RSPM on the basis of this observed pattern of neural activation, a more likely explanation here is that all novel and difficult tasks robustly activate domain- and process-general areas of the brain such as the WM core network. In fact, recent work by Fedorenko, Duncan, and Kanwisher (2013) has carefully identified a network which shows increased activation for hard > easy contrasts across a wide range of cognitively demanding tasks, including: precentral gyrus, anterior insula/frontal operculum, middle frontal gyrus, intraparietal sulcus, supplementary motor area, pre-supplementary motor area, anterior cingulate cortex, as well as posterior temporal and adjacent occipital regions. Interestingly, though perhaps not surprisingly, this primarily frontal and parietal network very closely resembles the core WM
network, the network associated with Jung & Haier’s (2007) P-FIT model of intelligence, as well as the regions evoked by harder > easier contrasts for all three tasks used in this study.

While a particular strength of the current study is the examination of both WM and $G_f$ functional neuroimaging tasks within the same population, several weaknesses may arise from their specific implementation. First, the dual n-back task was implemented in an ordered sequence from easiest (i.e. 1-back) to most difficult task blocks (i.e. 3-back), rather than a randomized order. While randomizing task block difficulty may have caused confusion in our participants, not doing so may have caused some ordering effects in which performance on latter difficulty levels was systematically influenced by earlier ones. Another limitation is the temporally imprecise nature of our block-design paradigm. For example, during the RSPM task, participants were given 8000 ms to decide whether the highlighted answer was correct. While some responses from participants did approach this timeframe, the average response time was in the range of 2000 ms to 3000 ms depending on the difficulty level. Thus, while the response window was required to be sufficiently long to account for all responses, current analyses potentially capture a large proportion of time unrelated to specific task completion. Much shorter response windows for the n-back and lexical decision tasks (2500 ms and 2000 ms respectively) mitigated this issue for these tasks. Our in-scanner RSPM task additionally failed to produce significantly different performance between difficulty levels for some metrics. For example, the medium and hard question sets were indistinguishable in terms of performance accuracy, and the easy and medium question sets were likewise indistinguishable in terms of neural activation patterns (despite statistical difference between all three levels of difficulty in terms of mean reaction time). These equivalencies suggest that splitting the original RSPM question set into two
rather than three levels of difficulty may have made for more distinct comparisons between difficulty levels.

Finally, the imprecision of using group level averages to comment on specific task activations should be acknowledged here. Inter-subject variability in neuroanatomy or true functional activation in response to a given task may contribute to overestimating task activation when averaged together in traditional group analysis (discussed in Nieto-Castañón & Fedorenko, 2012). This is particularly concerning in this case because we are acutely interested in functional activation overlap across two potentially exaggerated mappings. However, Fedorenko and colleagues (2013) comprehensively tested multiple-demand cortex via single subject analyses and found similar activation patterns to those resulting from traditional group level analyses, which provides some support for the validity of the current results. It is also worth mentioning here that with respect to the n-back task, different strategies for completing the task (e.g. exclusively relying on verbal rehearsal of auditorily presented letters, or visuospatial rehearsal of visual location stimuli) may result in differential patterns of activation regardless of ability level at any of the observed levels of difficulty. In other words, the n-back task poses a large enough cognitive problem that participants may focus variously on specific sub-tasks which may be reflected in variability in activation patterns in response to the task that are obscured at the level of the group average. Unfortunately, our current analyses are not able to exclude this possibility.

These limitations aside, the current study adds to the present understanding of the neural bases of WM and Gf in several important ways, particularly in the context of potential clinical, educational, or remedial applications such as WM training. First, the demonstrable overlap observed in dual n-back and RSPM tasks provides preliminary support for the neural basis of WM training, in that dual n-back and RSPM tasks do indeed activate similar areas of cortex
associated with task-specific and goal-directed behaviour. Secondly however, the large functional activation overlap between the supposed training task and a wide variety of other cognitively demanding tasks (e.g. our comparison lexical decision task) presents a challenge to the notion that overlapping activation alone is sufficient to account for any observed behavioural effects resulting from training. Future research can add clarity to the relationship between WM and $G_f$ by investigating overlapping activations with a more diverse set of representative tasks in each domain, as well as with a more diverse participant or patient populations (e.g. elderly populations, brain injured patients, or persons with schizophrenia). With regard to the WM training and its feasibility on the basis of overlapping neural activations for training and transfer tasks, future research should seek to gain a refined understanding of the conditions under which overlapping task activations ought to have behaviourally transferrable effects, and when they ought not to.
2.5 References


Chuderski, A. (2013). When are fluid intelligence and working memory isomorphic and when are they not? *Intelligence, 41*(4), 244-262. doi:10.1016/j.intell.2013.04.003


Chapter Three: Working Memory Training and Transfer to Untrained Cognitive Abilities: Support for the Null

3.1 Background

Working memory (WM) is the set of cognitive processes that work to maintain and manipulate task-relevant information during cognitive task performance, while also preventing interference from task-irrelevant information. In this sense, WM is an interplay between attention and memory that allows for temporary access to intermediate mental representations needed for more complex cognition. By briefly preserving task-relevant information, and facilitating manipulation of it, WM allows us to act outside the bounds of the immediate moment, and to coordinate complex and goal-directed behaviours (Baddeley, 1992; Repovs & Baddeley, 2006). As such, WM is a core cognitive ability in humans, and underlies performance on virtually all complex cognitive tasks, both within and beyond the laboratory. People differ in terms of how much information they can store in WM, and also in how readily they can store this information in the face of distraction (Engle, Kane, & Tuholski, 1999). While the absolute value of these inter-individual differences in WM capacity may in fact be quite small (e.g. 2 versus 6 items for low- and high-ability individuals respectively; Cowan, 2001), these differences have been found to be highly predictive of performance on a wide variety of cognitively demanding tasks, including: reading comprehension, language abilities, mathematics, reasoning, problem solving, and also overall academic performance (Engle, Tuholski, Laughlin, & Conway, 1999; Jaeggi, Buschkuehl, Shah, & Jonides, 2014).

In addition to driving variation in scholastic achievement and educational success, WM ability has also been found to be highly related to the ability to acquire knowledge, to learn new skills, and also to the construct of ‘fluid intelligence’ more broadly (Eriksson, Vogel, Lansner,
Bergström, & Nyberg, 2015). In the theory of Cattell (1963), ‘fluid intelligence’ ($G_f$) is the ability to adapt our reasoning abilities to solve novel cognitive problems. In contrast, ‘crystallized intelligence’ ($G_c$) draws heavily upon previously learned culturally-rooted knowledge acquired from education and previous experience (Carroll, 1996; McGrew, 2009; Carpenter, Just, & Shell, 1990). Fluid intelligence and WM are highly related psychological constructs. Working memory capacity has been established as one of the best predictors of general intelligence (Conway, Kane, & Engle, 2003), and investigations of the strength of the relationship between WM and $G_f$ in particular have indicated moderate correlations with coefficients in the .3 to .9 range (see Colom, Rebollo, Palacios, Juan-Espinosa, & Kyllonen, 2004; Burgess, Gray, Conway, & Braver, 2011). Similarly, Martínez and colleagues (2011) describe WM capacity and $G_f$ as almost isomorphic, and Chuderski (2013) noted latent factors of the two constructs being statistically indistinguishable when time limits were imposed on test takers. General intelligence itself, perhaps unsurprisingly, has been linked to a wide variety of important life outcomes, including academic success (Neisser et al., 1996; Watkins, Lei, & Canivez, 2007), job performance (Schmidt & Hunter, 1998), income (Ceci & Williams, 1997; Strenze, 2007), health (Gottfredson, 2004; Gottfredson & Deary, 2004), morbidity (Whalley & Deary, 2001), mortality (O’Toole & Stankov, 1992; Whalley & Deary, 2001), and crime (Neisser et al., 1996).

Given the strong relationship between WM and $G_f$, and the wide range of social, educational, and occupational outcomes to which they are positively correlated, it is no surprise that recent research has intensely focused on developing interventions to increase them via training (Jaeggi et al., 2014; Sprenger et al., 2013). Lövdén, Bäckman, Lindenberger, Schaefer, & Schmiedek (2010) discuss a theoretical framework for improving cognitive abilities in terms
of cognitive plasticity, noting that it can occur under conditions of a “prolonged mismatch between functional organismic supplies and environmental demands” (pp. 659). In other words, cognitive plasticity (or increases in WM ability, in this case), can occur when individuals are challenged only slightly above their current levels of functioning (see also von Bastian & Eschen, 2016). Halford, Cowan, & Andrews (2007) posited a model by which facilitation of one cognitive ability might then transfer to a different untrained ability. Specifically, they argued that Gf and WM are related in that both share a common capacity constraint due to a shared demand for attention in respective reasoning or memory tasks. Under this model, while a common capacity limit may be expressed in terms of the number of items a person is able to hold in WM, the same capacity limitation may be expressed in terms of the number of interrelations amongst elements a person is able to maintain during a reasoning task indicative of Gf ability. The general idea is that if working memory capacity could be increased, even just marginally by training, performance on other cognitive abilities that are strongly related to it (like Gf) ought to thereby be augmented as well. Working within this framework, Jaeggi and colleagues (2008) posited that training of WM specifically ought to facilitate Gf by expanding the common capacity restraint shared by both cognitive domains. Importantly, under this conceptualization, tasks that train WM capacity ought to facilitate Gf without simply teaching or practicing tests of Gf. Stated differently, facilitation of WM performance through training ought to generalize, or transfer, to performance of Gf.

Jaeggi and colleagues (2008) put the malleability of WM capacity to the test by training 34 participants across four groups with an adaptive dual n-back (i.e. requiring participants to track both visual and auditory stimuli from n trials previous) for 8, 12, 17 or 19 days. All groups were assessed before and after training with tests of Gf, and demonstrated higher scores on these
measures at post-test, following a positive dose-dependent relationship with time spent training (i.e., more WM training led to greater Gf gains). Essentially, Jaeggi and colleagues (2008) were able to demonstrate what decades of previous research had not: transfer of training gains from a specific cognitive task, to a more general intellectual capacity. This is what the literature has termed _far-transfer_, or the facilitation of performance on a task substantially different from a training task (Barnett & Ceci, 2002). In contrast, _near-transfer_ represents an expectable cognitive facilitation of performance on a task closely related to a given training task.

This initial study by Jaeggi and colleagues (2008), and subsequent positive findings in WM training (Chein & Morrison, 2010, Schmiedek, Loveden, & Lindenberger, 2010; Jaeggi, Studer-Luethi, Buschkuehl, Su, Jonides & Perrig, 2010; Jaeggi, Buschkuehl, Jonides & Shah, 2011; Schweizer, Grahn, Hampshire, Mobbs, & Dalgleish, 2013; Buschkuehl, Hernandez-Garcia, Jaeggi, Bernard, & Jonides, 2014; Hardy, Nelson, Thomason, Sternberg, Katovich, Farazin, & Scanlon, 2015) challenge the traditional view that cognitive abilities (versus acquired skills) reflect stable individual traits which are essentially fixed by early adulthood (Neisser, 1996). Rather, the success of training to induce far-transfer from WM capacity to Gf, suggests instead that the neural systems underlying WM do in fact remain plastic throughout the lifespan, and can be enhanced via intensive cognitive training.

Empirical study on WM training and its effects on Gf has greatly intensified since the publication of Jaeggi et al.’s (2008) findings, and although many studies have found strong and durable effects (over several months) for near-transfer of WM abilities, examples of far-transfer to Gf have been more elusive, as well as generally weaker and less durable when they have been found (see Colom et al., 2010; Owen et al., 2010; Chooi & Thompson, 2012; Redick et al., 2013, Thompson et al., 2013; Sprenger et al., 2013; Harrison, Shipstead, Hicks, Hambrick, Redick, &
Engle, 2013; von Bastian & Oberauer, 2013a; Colom et al., 2013, von Bastian & Eschen, 2016; Lawlor-Savage & Goghari, 2016). Rather, to this point there exists a striking lack of consensus in the literature about whether or not training on WM tasks generalizes to \( Gf \), and secondly, the specific methodology by which these claims ought to be tested. The topic remains highly controversial and has spurred a variety of conflicting reviews (Buschkuehl & Jaeggi, 2010; Conway & Getz, 2010; Takeuchi, Taki, & Kawashima, 2010; Morrison & Chein, 2011; Shipstead, Redick, & Engle, 2012; von Bastian & Oberauer, 2013b; Redick, Shipstead, Wiemers, Melby-Lervåg, & Hulme, 2015), meta-analyses (Melby-Lervåg & Hulme, 2013; Au, Sheehan, Tsai, Duncan, Buschkuehl, & Jaeggi, 2014; Schwaighofer, Fischer, & Buhner, 2015; Weicker, Villringer, & Thöne-Otto, 2016), meta-analytic rebuttals (Melby-Lervåg & Hulme, 2015), meta-analytic counter-rebuttals (Au, Buschkuehl, Duncan, & Jaeggi, 2015), and even further meta-analytic rejoinders (Melby-Lervåg, Redick, & Hulme, 2016) on the basis of existing trials. The resulting literature on the efficacy of WM training is what Urbánek and Marček (2016) have candidly called “reliably ambiguous” in terms of efficacy. Unfortunately the cumulative effect of this literature has been to jointly obfuscate the ostensibly simple question that each individual experiment, review, and meta-analysis has sought to clarify: “Does working memory training work?”

This ambiguity of efficacy for WM training has most often been discussed in terms of (and attributed to) shortcomings in study methodology throughout the literature dating back to Jaeggi and colleagues’ seminal experimental trial in 2008. While their positive findings spurred cautious optimism for cognitive enhancement through WM training, critics pointed to the necessity to tighten experimental controls and increase methodological rigor moving forward. Their provocative results led Sternberg (2008) to proclaim that “Increasing fluid intelligence is
possible after all”, and that “Fluid intelligence is trainable to a significant and meaningful degree” (pp. 6791). However, Sternberg (2008) also noted several shortcomings in Jaeggi et al.’s (2008) methodology and listed several associated open empirical questions regarding WM training and transfer to $G_f$ in order to guide future research. These included 1) the need for testing additional WM training tasks, and additional measures of $G_f$; 2) the need to demonstrate that gains in $G_f$ further generalize to the domains which $G_f$ is supposed to predict (i.e., educational, occupational attainment), and that those associations remain strong after $G_f$ has been increased via WM training; 3) the need to investigate the durability of training gains over time; 4) the need to implement a stronger control group as a comparison; 5) the need to investigate these effects in a broader range of participants; and, 6) the need for replication.

Subsequent investigations and reviews have addressed each of these points (as examples, see [1] von Bastian & Oberauer, 2013a; [2] Redick, et al., 2015; and [3-6] Melby-Lervåg & Hulme, 2013; Melby-Lervåg et al., 2016), however, new and more specific methodological qualms have arisen in the literature in an attempt to further homogenize study design, and encourage the search for additional unmeasured or uncontrolled variables which may account for significant variance in extant WM training trials. The search for these variables can generally be divided into two main types: 1) those relating to individual differences amongst WM training participants themselves; and 2) those relating to WM training trial design and execution.

Relating to individual differences amongst participants, Urbánek and Marček (2016) rightly point out, that from a conceptual point of view, the reliably ambiguous nature of the WM training literature may be the result of an (as of yet unmeasured) independent, randomly distributed factor in participants. For example, Chein and Morrison (2011) noted that no study up to that date had accounted for the potential effects of motivation, commitment, or training task
difficulty across experimental and control conditions. Jaeggi and colleagues (2014) echoed these concerns, and further suggested that individual differences in personality factors, pre-existing ability, and intrinsic versus extrinsic motivational factors need to be considered when assessing WM training and transfer.

Relatedly, recent work by Unsworth (2014) demonstrated that individual variance in three specific cognitive domains were necessary to fully mediate the relationship between WM capacity and *Gf* by training of WM: a) primary memory (i.e. the number of items that can be held in memory); b) secondary memory (i.e. the ability to strategically search for items currently being held in memory); and, c) attentional control (i.e. the ability to actively control attention in the service of current goals). Individuals vary in each of these abilities, and Unsworth (2014) suggests that the effectiveness of any proposed WM training paradigm may succeed or fail in transferring to *Gf* to the extent that all three abilities are engaged by the prescribed training task(s). Finally, and in response to earlier calls to examine non-ability based individual traits as potential sources of variance in WM training trials, Urbánek and Marček (2016) found that individual personality style was predictive of training gains associated with different methods of training. Specifically, and consistent with the few existing studies that have examined personality traits in relation to WM training (Studer-Luethi, Jaeggi, Buschkuehl, & Perrig, 2012; Savage, 2013), anxious, neurotic, critical, reserved, and cautious personality styles were associated with greater training gain from simpler WM training tasks. Conversely, participants with more intuitive and action-oriented personality styles showed greater proclivity for more difficult training tasks which generally place greater demands on more immediate processing by presenting diverse multi-modal stimuli (e.g. a triple n-back task).
Relating to WM training trial design and execution, Redick and colleagues (2015) discuss several methodological issues ubiquitous in the WM training literature as a type of ‘best practices guide’ to study design. Firstly, they advocate for the use of sensible active control groups over simple no-contact control groups. When compared to no-contact control groups alone, active training groups may benefit from a number of advantages related to the placebo or Hawthorne effects, including amount of contact with the experimenters or computers, familiarity or level of comfort with the research team or setting, as well as increased expectancy of gains or motivation to achieve them. Secondly, they stress the importance of adequate sample sizes, and recommend at least 20 participants per group, following Simmons, Nelson, & Simonsohn (2011). Small sample sizes are unfortunately common in the working memory training literature likely due to the time and cost associated with the intervention, and can produce inflated effect sizes. Third, if facilitation of $G_f$ by WM training is to occur by increasing the capacity of WM (as per Halford et al.’s 2007 model), evidence of this intermediate step should also be demonstrated along with evidence of the far-transfer by a separate task from the training task itself. Fourth, the pattern of results supporting the transfer effect should be ‘sensible’. That is, further than simply achieving a significant group by time interaction effect, this result should be achieved within the context of relatively equal group performance at pre-training testing, and divergent performance at post-training in favour of the active training group. Redick (2015) outlines a number of papers which report significant interaction effects via some other nonsensical pattern of group by time interaction (e.g. divergence at pre-training testing, and equivalence at post-training testing). Finally, Redick and colleagues (2015) advocate for including more than one outcome measure for far-transfer to $G_f$ which can then be used to form a composite or latent variable for subsequent analyses.
Meta-analytic work (Melby-Lervåg & Hulme, 2013; Au et al., 2014; Karbach & Verhaeghen, 2014, Weicker et al., 2016; Melby-Lervåg et al., 2016) has pointed to a number of potentially moderating factors of WM training trial success or failure, including type of cognitive training type (n-back training versus other types), participant age (younger versus older), participant status (learning disabled versus impaired WM versus normal functioning), training dose (less versus more), randomization (randomized versus nonrandomized), type of control group (treated versus untreated), geographic location (United States versus international populations), remuneration for participation (more versus less), and publication type (theses, dissertations, and conference posters versus journal articles, book chapters, and peer-reviewed conference proceedings). Unfortunately, as alluded to above, the authors of these meta-analytic reviews have disagreed about the appropriate methods for conducting a meta-analytic review of WM training, which have led them to opposite conclusions about the efficacy of WM training overall. Specifically, Melby-Lervåg and colleagues (2015, 2016), have charged Au and colleagues (2014, 2015) with: a) failing to account for baseline differences in training versus control groups while calculating effect sizes; b) including studies that rely on untreated control groups; 3) merging disparate outcome measures (e.g. reading comprehension and executive function) into a single far-transfer construct, and; 4) restricting trial selection to only one alternative of the above-listed potentially moderating variables, such as WM training type, participant age, or participant status, thereby limiting the scope of possible moderator analyses. Dougherty, Hamovitz, and Tidwell (2016) additionally critique Au and colleagues’ (2015) meta-analysis on the basis of observing a nearly seven times larger average effect size for studies utilizing untreated versus treated control groups ($g = 0.44$ versus $g = 0.06$), though still maintaining that control group type was not a significant moderator of treatment effect.
Melby-Lervåg et al.’s latest (2016) meta-analytic review addressed several of these shortcomings in examining 87 publications with 145 separate experimental comparisons of WM training groups versus treated control groups. Most germane to the present investigation, the authors did find a significant effect of for cognitive training nonverbal ability in adults ($g = 0.10; p < .05$), and for n-back training specifically ($g = 0.15; p < .05$) in studies using treated controls (effect sizes jump to 0.20 and 0.26 respectively when examining studies comparing to untreated controls). However, closer examination of the studies that contributed to this significant positive effect size were found to suffer from several of the methodological shortcomings described by Redick and colleagues (2015). For example, the five largest effect sizes were arrived at with sample sizes of less than 20 per group, and employed only a single outcome measure of nonverbal ability. More troublingly, four of these five largest effect sizes evinced substantial unexplained decreases in outcome measure scores for the control group, which were in fact larger than the increases observed in the training groups. These nonsensical (or at least conceptually counterintuitive) ‘crossover patterns’ of training effect (Redick, 2015; Wagenmakers, 2015) artificially inflate the effect sizes for individual comparisons, as well as for averaged estimates in meta-analyses. Melby-Lervåg and colleagues (2016) additionally note that the effect size of n-back training on nonverbal ability drops below significance with only the most problematic of these five studies removed from the analysis. Perhaps most troublingly of all, observed gains in nonverbal ability were not found to be significantly related to increases in WM abilities themselves, thereby casting doubt on the proposed mechanism of far-transfer discussed by Halford et al. (2007). Overall, Melby-Lervåg and colleagues (2016) conclude that while there is convincing evidence of large improvements on tasks similar to those utilized by WM training (i.e. near-transfer, and ‘intermediate transfer’ to visual and verbal WM), there are
no convincing effects of far-transfer of WM training to constructs such as nonverbal ability, verbal ability, reading comprehension or arithmetic that could not otherwise be explained by methodological shortcomings. Importantly, and contrary to the suggestions in the literature regarding potential effects of individual differences, moderator analyses revealed no evidence of moderation effects for nonverbal ability (e.g. participant age, status, training dose, training type etc.) aside from significantly higher effect sizes for studies utilizing untreated controls versus those implementing treated control groups. Crucially, Melby-Lervåg and colleagues (2016) demonstrated the effect of adequate sample size and control group treatment by pooling effect sizes for studies falling into the four resulting permutations of study design (i.e. ≥ 20 participants and treated controls, ≥ 20 participants and untreated controls, < 20 participants and treated controls; < 20 participants and untreated controls). Average effect sizes in each of these conditions showed significant effects for far-transfer of WM training to nonverbal ability, except for the most conservative and robust experimental design (i.e. ≥ 20 participants and a treated control group), which showed an average effect size close to zero (g = 0.01).

Finally, recent work has suggested that beyond the methodological variables discussed above, results of WM training studies may in fact be substantially influenced by how participants self-select for participation before the experimental trial even begins. In a unique experimental design, Foroughi, Monfort, Paczynski, McKnight, & Greenwood (2016) reported differential effects of suggestive versus non suggestive recruitment materials. Specifically, participants who responded to recruitment posters which overtly listed (or suggested, or implied) benefits of cognitive training (e.g. “Brain Training & Cognitive Enhancement”), outperformed participants who responded to recruitment posters that did not overtly list or suggest such benefits (e.g. “Email Today & Participate in a Study”) on measures of nonverbal reasoning after only a single
hour of cognitive training. The authors note that the magnitude of increase on these measures equates to an approximate five to ten point increase on a standardized 100-point intelligence scale. Rather than attributing the enhanced performance to true gains in intelligence from the single hour of cognitive training, the authors discuss their results in terms of a placebo effect from suggestive recruitment materials, as well as a self-selection bias in this placebo group for those who have stronger beliefs in the malleability of intelligence in general. Foroughi and colleagues (2016) stress the importance of this finding to current efforts in the literature to clarify the efficacy of WM training, essentially introducing specific recruitment methodology as another potential (and largely unmeasured and/or unreported) moderator of treatment effect in previous trials. The authors suggest two potential solutions to the problem for future research: 1) recruit participants for WM training studies covertly, sharing as few details regarding intended comparisons as possible (Boot, Simmons, Stothart, & Stutts, 2013; Shipstead et al., 2012), or; 2) recruit only those participants who believe the treatment will work, and entrust appropriate methodological controls to separate true effects from participant expectation and motivation. This later view is consistent with suggestions from the psychotherapy literature which stresses the importance of motivation to treatment effectiveness (Ryan, Lynch, Vansteenkiste, & Deci, 2010), as well as indications from the medical literature suggesting that expectations of participants can be better utilized to test treatment effects in randomized controlled trials (Torgerson, Klaber-Moffett, & Russell, 1996).

Dougherty and colleagues (2016) summarize the impact of these methodological issues on the WM training literature succinctly, noting that the particular choice of experimental design ought not to moderate the effectiveness of a manipulation, except in cases where the experimental design itself is responsible for creating the effect through confounding variables.
They suggest that this is in fact currently the case for successful WM trials with untreated control conditions, where participant expectations are confounded with engaging in training, ultimately leading to placebo effects in ostensibly ‘trained’ participants.

Given this cursory overview of the rapidly expanding and evolving field of WM training, the present study seeks to address whether or not the pattern of far-transfer of ability from WM capacity to Gf can be replicated while addressing several of the methodological shortcomings ubiquitous to the current literature. The most up to date meta-analytic review of the field at the time of planning the current study was that of Melby-Lervåg and Hulme (2012), which included results from 30 comparisons from 23 studies carried out between 2002 and 2011. While more recent reviews (discussed above) have become increasingly pessimistic about true effects of WM training, they also have the benefit of drawing from a pool of experimental investigations almost five times as large as that of Melby-Lervåg and Hulme’s initial work in 2013, just three years later (recall that Melby-Lervåg et al.’s latest (2016) meta-analytic review includes 145 comparisons from 87 separate studies). Thus, while the most up to date reviews tend to support the null hypothesis, earlier reviews were somewhat more optimistic – and particularly so for n-back training in young adults transferring to nonverbal abilities. For instance, Melby-Lervåg and Hulme (2012), reported moderate effect sizes of 0.37 for studies examining training in young adults, and 0.34 for studies utilizing the n-back task specifically.

On the basis of these early initial estimates of effect size in the literature, we hypothesized that: 1) WM trained participants would demonstrate increased task performance on the training tasks themselves, 2) as well as increased WM capacity (i.e. near-transfer), compared to our treated and untreated control comparison groups. We additionally hypothesized that: 3)
participants in the WM training group would exhibit far-transfer of ability to untrained tasks via increased test scores on measures of Gf compared to the treated and untreated control groups.

3.2 Method

3.2.1 Participants and Recruitment

A total of 359 healthy adults were recruited via printed advertisements distributed throughout the community as part of a larger neuroimaging WM training trial. Participants completed online screening measures at braintrainingstudy.ca which inquired about study exclusion criteria, including: 1) age less than 18, or greater than 40; 2) left-handedness; 3) history of traumatic brain injury or other neurological condition causing sensory or motor impairment; 4) self-reported presence of Axis I mental illness; 5) less than normal or corrected-to-normal visual acuity; 6) MRI contra-indications; 7) insufficient access to a computer and high-speed internet; and 8) recent or previous use of the n-back training task or other online cognitive training paradigms. Of the 359 participants who completed the screening questionnaires, 187 were invited to participate in the study, and a total of 76 participants were ultimately included in the analyses. See Figure 8 for a flow chart depicting the recruitment, randomization, and exclusion process.
Figure 8. Flow chart of study design. *Two participants in MRI conditions were reassigned to the no-contact control group after being unable to tolerate MRI scanning. †Participants removed from analysis due to training contamination, low training dosage, or data acquisition issues.
Participants were compensated $20 per cognitive testing session, and $20 per MRI session, totalling $80 for the four appointments attended by participants randomly assigned to the MRI conditions, and $40 for the two appointments attended by those assigned to the no-contact control condition. Written consent was obtained from all participants, and ethics approval was obtained from the University of Calgary’s Conjoint Health Research Ethics Board (CHREB).

3.2.2 Procedure and Materials

Following initial recruitment and screening, participants were randomized to one of three groups: a WM training group \((n = 25)\), a processing speed (PS) active control group \((n = 24)\), or a no-contact control group \((n = 27)\). PS training was chosen as an active control condition on the basis of its association with robust improvements on measures of processing speed, but not measures of WM, inhibition, or nonverbal reasoning (Takeuchi & Kawashima, 2012). Thus, preliminary evidence suggests that PS training may offer a viable active control condition to WM training by holding constant the level of effort, motivation, interaction with computers and researchers for participants, while impacting relatively orthogonal behavioral skillsets (Takeuchi et al., 2011; Takeuchi & Kawashima, 2012).

Participants were blinded to group randomization with respect to the WM and PS training groups. However, because assignment to the no-contact control group entailed not undergoing MRI scanning sessions, and not completing online training, participants in this group were aware of their group assignment. Efforts were made to blind experimenters to group assignment, though the distinction between training groups versus no-contact control was similarly difficult to blind because of the difference in the number of scheduled appointments (i.e. two additional MRI appointments for participants in the WM and PS training groups). In this sense, the experimenters cannot be considered to have been truly blind to group assignment. Importantly
however, the experimenters were typically unable to distinguish between those in the WM training versus PS active control groups when meeting them for MRI or cognitive testing appointments. Following group assignment, participants in the WM and PS training groups underwent their initial MRI session, and then completed initial cognitive testing appointments on a separate day shortly thereafter. They were then given login access to Lumosity.com (Lumos Labs Inc., 2009) where they were asked to complete specially designed online training programs targeting either WM or PS cognitive processes. Participants were asked to complete at least 20 minutes of training per day, for five out of seven days per week, for six weeks. Progress in training was monitored online for each participant, and individuals were removed from the study if they did not complete at least 20 of the assigned 30 days of cognitive training over the six week training period. Participants were also removed from the study if they erroneously accessed Lumosity training games outside of those prescribed by their training program. Encouragement emails were sent to participants on a weekly basis in order to facilitate compliance with the prescribed training regimen. Following training, participants in the WM training and PS active control groups underwent a second cognitive assessment. Participants in the no-contact control group simply completed cognitive testing on two occasions, approximately six weeks apart.

3.2.3 Cognitive Testing and Behavioural Measures

Cognitive testing included split-half subtests from the Wechsler Adult Intelligence Scale – Fourth Edition (WAIS-IV; Wechsler, 2008), Raven’s Advanced Progressive Matrices (RAPM; Raven, 1975; Raven & Raven, 1994), and Cattell’s Culture Fair Test (CCFT; Cattell & Cattell, 1959, 1973). Importantly, parallel forms (i.e. split halves) of these cognitive measures were not randomized across pre- versus post-training assessments, though order of administration was pseudorandomized. Thus, participants in all groups completed odd numbered items before
training, and even numbered items after, in the same pseudorandomized order across both testing sessions. Computerized administrations of the Automated Operation Span Task (AOSPAN; Unsworth, Heitz, Schrock, & Engle, 2005), and a Spatial Delayed Response Task (SDRT; Glahn et al., 2002) were also administered both before and after training. Cognitive assessments were completed by PhD-level graduate students with specific training in neuropsychological assessment, or trained undergraduate volunteers. Assessment sessions were typically 100 to 120 minutes in duration. Psychometric properties listed in test manuals are discussed for each test below, and test-retest reliability values as calculated based on our no contact control group (n = 27) are listed in Table 2.

3.2.3.1 Wechsler Adult Intelligence Scale – Fourth Edition (WAIS-IV)

Eight of the 10 core subtests of the WAIS-IV were administered in order to allow calculation of all four composite indices of intelligence assessed by the WAIS-IV: Verbal Comprehension Index (VCI), Perceptual Reasoning Index (PRI), Working Memory Index (WMI), and Processing Speed Index (PSI). These included ‘Vocabulary’, ‘Similarities’, ‘Block Design’, ‘Matrix Reasoning’, ‘Digit Span’, ‘Arithmetic’, ‘Symbol Search’, and ‘Coding’. All subtests were split in half for pre- versus post-training comparison, with the exception of Digit Span, Symbol Search, and Coding, which were administered in their entirety before and after training. Discontinue rules for split-half subtests were halved and rounded up where necessary. Importantly, all subtests of the WAIS-IV demonstrate excellent psychometric properties with good evidence of internal consistency, test-retest reliability, and interscorer agreement. Across the 13 age groups for which psychometric properties are provided by the WAIS-IV manual, internal consistency (split-half) estimates ranged from .97-.98 for the full-scale intelligence quotient (FSIQ), .87-.98 for Index scores, and .71-.96 for individual subtests. Similarly, short
term test-retest stability (tested with an average delay of 22 days) was found to be highest for FSIQ, followed by VCI, PRI, WMI, PSI, and lower for each individual subtest.

3.2.3.2 Raven’s Advanced Progressive Matrices (RAPM)

RAPM (Raven, 1975; Raven and Raven 1994) is a reliable and well validated paper and pencil test of general cognitive ability. Participants are asked to examine a matrix pattern with a missing piece, and select the correct answer from eight possible answers. RAPM is published in two sets: Set-I which contains 12 screener and/or practice items and has a five minute time limit, and Set-II which contains 36 items and has a 40-minute time limit. The reliability and validity of the RAPM has been studied for decades, however most recently, NCS Pearson (2011) reported an internal consistency (split-half) of .85, and strong convergent validity with other measures of general intellectual ability. Importantly, the RAPM is considered to be at the centre of a constellation of tests that measure analytic or fluid intelligence (Carpenter, Just, & Shell, 1990), making it a premier test of functioning in this domain (Alderton & Larson, 1990).

3.2.3.3 Cattell’s Culture Fair Test (CCFT)

CCFT (Cattell & Cattell, 1959, 1973) is a test of general reasoning and cognitive abilities that was designed specifically to reduce emphasis placed on linguistic abilities and general store of culturally specific knowledge in traditional tests of intelligence. The test contains two equivalent forms, each consisting of four subtests: series, analogies, matrices, and classification, and thus provides a more varied assessment of general cognitive functioning beyond matrix reasoning ability as assessed in isolation by the RAPM (Colom & Garcia-Lopez, 2003). Like the RAPM, CCFT has demonstrated good psychometric properties, with an internal reliability of .85, test-retest reliability of .82, and a high degree of validity in relation to other measures of general
intelligence (Cattell & Cattell, 1959). For all participants, form A was administered at pre-training cognitive testing, and form B was administered at post-training testing.

3.2.3.4 Automated Operation Span Task (AOSPAN)

The AOSPAN task (Unsworth et al., 2005) is a complex measure of WM which requires participants to remember the sequential ordering of presented stimuli while carrying out simple mathematic problems as a distraction. The task has been used previously in the WM training literature (Jaeggi et al., 2010) and has demonstrated good psychometric properties, including test-retest reliability of .83, and good concordance with other measures of WM (Unsworth et al., 2005). Several outcome scores are tabulated by the task including a ‘score’ representing the sum of all perfectly recalled sets of letters, and a ‘total’ representing the number of correctly recalled letters with respect to serial position regardless of whether the entire set was recalled correctly. Here, we report on the latter, as it is theoretically a more sensitive indicator of change in WM ability in response to training.

3.2.3.5 Spatial Delayed Response Task (SDRT)

The SDRT (Glahn et al., 2002) assesses visuospatial working memory by briefly presenting participants with a series of circles on a computer screen, and requires that they determine whether a second configuration of circles is the same after a two second delay. A second condition asks participants to determine whether the second set of circles is the same as the first set, but flipped about the horizontal midline of the presentation space. Across a variety of difficulties (1, 3, 5, or 7 circles presented), the variable of interest is the total number of correct trials for both with- and without-manipulation (i.e., flipped) conditions.
Table 2.

Test-retest reliability correlations for cognitive tests, based on scores from the no-contact control group

<table>
<thead>
<tr>
<th>Test Re-test Reliability Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>WAIS-IV</strong></td>
</tr>
<tr>
<td>Full-Scale IQ</td>
</tr>
<tr>
<td>Verbal Comprehension Index</td>
</tr>
<tr>
<td>Vocabulary</td>
</tr>
<tr>
<td>Similarities</td>
</tr>
<tr>
<td>Perceptual Reasoning Index</td>
</tr>
<tr>
<td>Block Design</td>
</tr>
<tr>
<td>Matrix Reasoning</td>
</tr>
<tr>
<td>Working Memory Index</td>
</tr>
<tr>
<td>Digit Span</td>
</tr>
<tr>
<td>Arithmetic</td>
</tr>
<tr>
<td>Processing Speed Index</td>
</tr>
<tr>
<td>Symbol Search</td>
</tr>
<tr>
<td>Coding</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Computerized Working Memory Tasks</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Automated Operation Span</td>
</tr>
<tr>
<td>SDRT Spatial Maintenance Task</td>
</tr>
<tr>
<td>SDRT Spatial Maintenance and Manipulation Task</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Nonverbal Reasoning Tasks</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Raven’s Advanced Progressive Matrices</td>
</tr>
<tr>
<td>Cattell’s Culture Fair Test</td>
</tr>
</tbody>
</table>

3.2.3.6 Additional Behavioural Measures

In addition to the above cognitive assessments, participants were also asked to complete questionnaires on a wide variety of other characteristics which might mediate or moderate observed effects of online cognitive training. These included measures of personality (HEXACO; Ashton & Lee, 2009), need for cognition (Cacioppo, Petty, & Kao, 1984), ‘grit’ (i.e. commitment to long term goals; Duckworth & Quinn, 2009), and current cognitive activities (Eskes, et al., 2010). Participants in the WM training and PS active control groups were also asked to complete training-specific measures of motivation to complete training, and
expectations of cognitive improvement as a result of training. Measuring motivation and expected benefits of training is particularly important given the literature regarding the potential for motivational factors to artificially facilitate training effects (see Boot et al., 2013; Foroughi et al., 2016). Participants in the no-contact control group did not complete any training, and were thus not assessed for motivation of expectancy effects. All questionnaires were administered once at the beginning of the study, with the exception of the motivation and expectancy questionnaire which was administered three times: before, mid-way through, and after training.

3.2.4 Training Tasks

3.2.4.1 Working Memory Training Program

Participants randomly assigned to the WM training group completed their online training with three games selected from Lumosity’s broader game library (Lumos Labs Inc., 2009) which specifically target WM processes: 1) ‘Memory Match’ is essentially a visual 2-back task which presents participants with an array of shapes progressing from right to left across the screen, advancing one position per trial. As the line of randomly ordered shapes progresses across the screen, it passes through two location indicator boxes two positions apart (i.e. one space between them). On each trial, participants are asked to indicate via button press whether the stimuli in the rightmost box matches that in the leftmost box which contains the stimuli from the rightmost box from two trials previous. This would be a simple matching task except that the shapes to the left of the first indicator box become invisible after several correct responses. This taxes memory for which shape was presented two trials previously, and requires continuous updating of the presented sequence. If participants respond incorrectly, all shapes in the sequence become visible until several subsequent correct responses render these shapes invisible again. 2) ‘Memory Match Overload’ is structured similarly to Memory Match, but leaves two spaces between
position indicator boxes, thereby making it a more difficult visual 3-back memory task. 3) Finally, ‘Memory Lane’ mimics the logic and cognitive challenge of the dual n-back task. Participants are guided down a virtual street in which each apartment building they pass acts as one trial of the dual n-back task. At each apartment, a human silhouette appears in one of the windows and auditorily presents a letter of the alphabet. Participants are instructed to indicate via button press if either or both the location of the silhouette in the window, and auditorily presented letter are the same as n apartments ago. Unlike the previous two training tasks, Memory Lane is adaptive in that the difficulty is increased when participants are successfully completing the task, and decreased when they are not, thereby ‘adapting’ to their skill level. The size of the visual stimuli presentation area (i.e. number of windows per apartment; 2x2 to 3x3), target n are (i.e. number of apartments ago to remember; 1-back to 10-back), and stimuli modality (i.e. visual only vs. both visual and audio) are adjusted accordingly. Screen captures of each of these tasks are shown in panel A of Figure 9. Each training day, participants would play each of these games several times in a randomized order. Participants were asked to engage in these tasks for 20 minutes per day, on at least five days per week, for six weeks.
Figure 9. Lumosity working memory training tasks (panel A) and processing speed active control tasks (panel B).

3.2.4.2 Processing Speed Training Program

Participants randomly assigned to the processing speed active control training group completed three different games from Lumosity’s broader game library (Lumos Labs Inc., 2009) that are heavily dependent on processing speed abilities: 1) ‘Speed Match’ is essentially a speeded visual 1-back task. It sequentially presents a series of shapes, and asks participants to quickly indicate via button press whether or not the present shape matches the one presented immediately before it. While this is a relatively simple task, emphasis is placed on improving
speed of responding over the course of training. 2) ‘Speed Match Overdrive’ shares a similar structure to Speed Match, but includes a third response option for the currently presented shape to be a ‘partial’ match to that presented directly before it (e.g. matches in colour but not shape, or matches in shape but not colour). Finally, 3) ‘Spatial Speed Match’ shares the same structure as Speed Match, but includes stimuli differing only in spatial orientation. For example, two empty dots and one filled dot might be shown followed by a similar arrangement with the filled dot in a different location. Importantly, these processing speed tasks were not directly adaptive in the way that the dual n-back training was made more or less difficult by altering variables of the game. However, there was an emphasis on constant improvement through reduction of reaction times over the course of training. Screen captures of each of these tasks are shown in panel B of Figure 9. Consistent with the WM training group, participants in the processing speed group would play each of these three games in random order for approximately 20 minutes on each of their training days.

3.2.5 Data Analysis

Potential differences between the three groups before training were investigated with one-way ANOVAs, chi-squared tests, or independent samples t-tests when comparing data pertaining only to the two active training groups. To determine whether training had precipitated significant changes in cognitive test scores, a mixed-design repeated measures ANOVA was undertaken, examining time (within-subjects; before training versus after training) × group assignment (between-subjects; WM training versus PS active control versus no-contact control group) for each of the cognitive tests in our pre- and post-training test battery. For all administered subtests of the WAIS-IV, scores were converted to age-appropriate scaled scores,
in order to calculate composite indices for verbal comprehension, perceptual reasoning, working memory, processing speed (VCI, PRI, WMI, PSI), as well as full-scale intelligence (FSIQ).

In addition to this traditional null hypothesis significance testing, Bayesian factors were calculated for each cognitive test via JZS Bayesian repeated measures ANOVAs in JASP version 0.8.0.0 for Windows (Love et al., 2015; JASP Team, 2016). JASP allows for the calculation of Bayes factors for a variety of different models, including the null hypothesis, each main factor individually (e.g. time or group), main factors combined (e.g. time + group), as well as the main factors combined with the interaction effect (e.g. time + group + time × group). Here we modelled each of the main factors as nuisance variables in order to include them with the null hypothesis, such that the interaction effect of interest (e.g. time × group) could be compared directly with its main explanatory rival – the null hypothesis including the main effects of time and group. In this way, Bayesian analyses, and Bayesian factors provide relative evidence of both null and alternative hypotheses, compared to the conclusions about the null hypothesis only, as proffered by traditional null hypothesis significance testing (Masson, 2011; Jarosz & Wiley, 2014; Rouder; Morey, Swagman, & Wagenmakers, 2016).

3.3 Results

3.3.1 Participant Demographics, Cognitive Characteristics, and Personality Variables

Participant groups were equivalent on all variables measured pertaining to demographics, and cognitive ability, including: age $[F(2,73) = 0.10, p = .90]$; balance of males and females $[\chi^2(2, N = 76) = 0.06, p = .97]$; years of education $[F(2,72) = 1.17, p = .32]$; estimated full-scale intelligence quotient $[F(2,73) = 0.70, p = .50]$, RAPM performance $[F(2,73) = 2.25, p = .11]$; CCFT performance $[F(2,73) = 1.62, p = .21]$; AOSPAN performance $[F(2,72) = 0.28, p = .76]$; and SDRT performance for both maintenance $[F(2,72) = 2.32, p = .11]$ and maintenance plus...
manipulation \[ F(2,72) = 1.85, p = .17 \] conditions. Groups were also equivalent on a number of scales measuring personality characteristics, including: the Grit scale \[ F(2,69) = 0.62, p = .54 \]; the Need for Cognition scale \[ F(2,70) = 0.52, p = .60 \]; the Cognitive Activities Questionnaire \[ F(2,65) = 2.22, p = .12 \], as well as all dimensions of the HEXACO personality inventory. The two training groups were also equal in terms of their self-rated motivation to complete training \[ t_{47} = -0.39, p = .70 \], and their expectations of improvement on the training tasks themselves \[ t_{47} = 0.35, p = .73 \]. Table 3 summarizes these results.
### Table 3.

**Participant characteristics**

<table>
<thead>
<tr>
<th>Demographics</th>
<th>Working Memory Training Group</th>
<th>Processing Speed Control Group</th>
<th>No-Contact Control Group</th>
<th>Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>25</td>
<td>24</td>
<td>27</td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>30.68 (6.24)</td>
<td>31.33 (5.78)</td>
<td>31.32 (5.58)</td>
<td>F(2,73) = 0.10, p = .90</td>
</tr>
<tr>
<td>Gender (male/female)</td>
<td>11/14</td>
<td>10/14</td>
<td>11/16</td>
<td>(\chi^2(2, N = 76) = 0.06, p = .97)</td>
</tr>
<tr>
<td>Education (years)</td>
<td>15.24 (2.19)</td>
<td>15.57 (1.93)</td>
<td>16.07 (1.84)</td>
<td>F(2,72) = 1.17, p = .32</td>
</tr>
</tbody>
</table>

**Cognitive Ability Before Training**

| WAIS-IV FSIQ     | 108.24 (15.93)                | 111.63 (12.34)                | 107.07 (13.81)           | F(2,73) = 0.70, p = .50 |
| RAPM             | 12.60 (3.01)                  | 13.50 (2.27)                  | 11.89 (2.76)             | F(2,73) = 2.25, p = .11 |
| CCFT             | 26.68 (4.43)                  | 28.46 (3.51)                  | 26.41 (4.93)             | F(2,73) = 1.62, p = .21 |
| AOSPAN total     | 52.79 (14.45)                 | 53.96 (12.69)                 | 50.93 (16.40)            | F(2,72) = 0.28, p = .76 |
| SDRT Maintenance | 17.56 (1.29)                  | 18.25 (1.39)                  | 17.35 (1.85)             | F(2,72) = 2.32, p = .11 |
| SDRT Manipulation| 16.04 (2.21)                  | 16.63 (2.10)                  | 15.38 (2.50)             | F(2,72) = 1.85, p = .17 |

**Personality Factors**

| Grit score       | 3.44 (0.69)                   | 3.31 (0.61)                   | 3.44 (0.45)              | F(2,70) = 0.36, p = .70 |
| Need for Cognition score | 69.36 (6.76)   | 69.06 (9.51)                  | 66.92 (10.60)            | F(2,70) = 0.52, p = .60 |
| Cognitive Activities (hours/year) | 1203.80 (890.50) | 1407.52 (1164.60)            | 816.28 (788.30)          | F(2,65) = 2.22, p = .12 |
| HEXACO Honesty-Humility | 3.74 (0.50)      | 3.80 (0.54)                   | 3.60 (0.60)              | F(2,70) = 0.92, p = .40 |
| HEXACO Emotionality | 3.00 (0.66)       | 2.83 (0.80)                   | 3.18 (0.60)              | F(2,70) = 1.53, p = .22 |
| HEXACO Extraversion | 3.37 (0.53)       | 3.54 (0.76)                   | 3.23 (0.59)              | F(2,70) = 1.42, p = .25 |
| HEXACO Agreeableness | 3.18 (0.54)      | 3.37 (0.50)                   | 3.20 (0.43)              | F(2,70) = 1.03, p = .36 |
| HEXACO Conscientiousness | 3.84 (0.51)     | 3.53 (0.64)                   | 3.68 (0.48)              | F(2,70) = 1.89, p = .16 |
| HEXACO Openness to Experience | 3.62 (0.67) | 3.93 (0.42)                   | 3.62 (0.63)              | F(2,70) = 2.23, p = .12 |

**Training Data**

| Total hours of training | 13.49 (4.86) | 11.69 (3.03) | - | \(t_{47} = 1.55, p = .13\) |
| Pre-training motivation | 5.68 (0.87) | 5.77 (0.78) | - | \(t_{47} = -0.39, p = .70\) |
| Pre-training expectation of improvement | 4.52 (0.96) | 4.56 (1.10) | - | \(t_{47} = 0.35, p = .73\) |

Note. FSIQ, full scale intelligence quotient; RAPM, Raven’s Advanced Progressive Matrices; CCFT, Cattell’s Culture Fair Test; AOSPAN, Automated Operation Span Task; SDRT, Spatial Delayed Response Task.
Behavioural Results

3.3.1.1 Training Task Performance and Reaction Time

Members of both the WM training group and the PS active control group showed improvement on their assigned training measures across the training period. Training progress was measured for each training game by calculating a difference score between their performance on their first game, and an average of their last five games. Participants in the WM training group achieved an average \( n \) of 1.80 (\( SD = 0.41 \)) on their first attempt of the Memory Lane game, and progressed to an average \( n \) of 4.79 (\( SD = 2.08 \)) across their last five games, yielding a significant average difference score of 2.99 (\( SD = 2.07 \)), \( t(24) = 7.24, p < .001 \). Additionally, WM training participants demonstrated increased proficiency on both the Memory Match, and Memory Match Overload games as indicated quantitatively by a greater number of correct matches across their last five games, compared to their first game. Difference scores were significant for both Memory Match \( t(24) = 10.15, p < .001 \), and Memory Match Overload \( t(24) = 12.41, p < .001 \). Participants also attempted a greater number of trials for these matching tasks over the course of training indicating quicker reaction times, and thereby being able to fit in a greater number of trials in a given 20-minute training session. Training progress in the processing speed active control group was indicated by significant decreases in reaction time across the training period. On average, participants reduced their reaction times by 367.63 ms (\( SD = 233.11 \)) on the Spatial Speed Match game, 278.50 ms (\( SD = 142.27 \)) on the Speed Match game, and 589.62 ms (\( SD = 228.14 \)) when comparing their first game to the average of their last five games. Difference scores indicated significant reductions in reaction time for each of these \( t(23) = 7.73 - 12.66, p < .001 \). These training results are displayed graphically in Figure 10.
In order to ensure that significant differences scores on these training tasks were not driven solely by the relatively large gains made in performance in the first several trials, the statistical tests were repeated with differences scores calculated with the first two trials removed – i.e. based on the difference between their performance or reaction time on their third trial versus that of the average of their last five. In this way, potential effects of simple familiarity with the computer/task interface can be at least partially removed. These analyses revealed the same pattern of results as the original analysis for both groups. Participants in the WM training group achieved an average $n$ of 1.96 ($SD = 0.20$) on their third attempt of the Memory Lane game, yielding a significant average difference score of 2.83 ($SD = 2.08$), $t(24) = 6.80$, $p < .001$. These difference scores were again significant for both Memory Match $t(24) = 7.74$, $p < .001$, and Memory Match Overload $t(24) = 10.35$, $p < .001$. Similarly, average reduction in reaction time remained significant for members of the processing speed training group when discarding the first two trials $t(23) = 7.81 - 13.16$, $p < .001$. 
Figure 10. Mean performance by training game for the working memory training group (A-C), and mean reaction times by training game for the processing speed training group (D-F).
Importantly, the training groups were observed to have spent a statistically equivalent amount of time training with their respective online training programs over the course of the roughly six week training period: 13.69 hours for the WM training group \((SD = 4.86)\), and 11.69 hours \((SD = 3.03)\) for the PS active control group; \(t(47) = 1.55, p = .13\).

3.3.1.2 Convergent Validity of Theoretically Related Cognitive Measures

Across all three groups, cognitive test scores obtained at the pre-training cognitive assessments showed the expected pattern of convergent validity within their respective broader cognitive domains. Measures of \(Gf\) (RAPM, CCFT total, WAIS-IV Matrix Reasoning, WAIS-IV Block Design; Benson, Hulac, & Kranzler, 2010) demonstrated Pearson correlation coefficients between 0.40 and 0.58 \((p’s < .001)\). Similarly, measures of WM (AOSPAN total, WAIS-IV Digit Span total, WAIS-IV Arithmetic) showed good coherence, though with some variability likely due to the difference in task presentation modality (i.e. verbal versus visual), and task demands (i.e. mathematic calculations versus simple memory span). WAIS-IV Digit Span total scores were significantly correlated with both WAIS-IV Arithmetic \((r = 0.32, p = .005)\), and AOSPAN \((r = 0.50, p < .001)\), as well as SDRT Maintenance and Manipulation \((r = 0.33, p = .004)\), but not SDRT Maintenance \((r = 0.10, p = .38)\). AOSPAN scores were not significantly correlated with WAIS-IV Arithmetic \((r = 0.07, p = .54)\), or SDRT Maintenance \((r = 0.13, p = .26)\), or Maintenance and Manipulation \((r = 0.18, p = .12)\). WAIS-IV Arithmetic was significantly correlated with SDRT Maintenance and Manipulation \((r = 0.27, p = .019)\), but not Maintenance alone \((r = 0.18, p = .12)\). Finally, the two WAIS-IV measures of processing speed (Symbol Search, Coding) were highly correlated \((r = 0.78, p < .001)\).
3.3.1.3 Cognitive Test Scores Before and After Training

Results of the mixed-design repeated measures ANOVA examining time × group for performance on cognitive testing revealed significant main effects of time for two age-normed indices of the WAIS-IV including PRI \[ F(1,73) = 24.41, p < .001 \], PSI \[ F(1,73) = 31.16, p < .001 \], as well as the AOSPAN task \[ F(1,72) = 11.85, p = .001 \], RAPM \[ F(1,73) = 4.86, p = .031 \], and CCFT \[ F(1,73) = 102.22, p < .001 \]. When raw scores from WAIS-IV subtests were used rather than age-normed composite index scores, main effects of time were found for Vocabulary \[ F(1, 73) = 13.41, p < .001 \], Similarities \[ F(1,73) = 6.57, p = .012 \], Block Design \[ F(1,73) = 37.70, p < .001 \], Symbol Search \[ F(1,73) = 12.16, p = 0.001 \], and Coding \[ F(1,73) = 31.35, p < .001 \]. See Figure 11.
Figure 11. WAIS-IV performance by group, before and after training. Verbal Comprehension Index (A); Perceptual Reasoning Index (B); Working Memory Index (C); Processing Speed Index (D); Full Scale Intelligence Quotient (E). Error bars represent 95% confidence intervals.
Additionally, the repeated measures ANOVA revealed main effects of group membership only for the SDRT spatial maintenance task \( F(2,72) = 3.96, p = .023 \), though very nearly for RAPM \( F(2,73) = 3.09, p = .051 \), and the Coding subtest of the WAIS-IV \( F(2,73) = 2.99, p = .057 \). Follow-up pairwise analyses using the Bonferroni correction revealed a significant difference only between the PS active control group (higher scores), and the no-contact control group (lower scores) for the SDRT maintenance task. This finding is corroborated by visual inspection of the obtained data for the SDRT maintenance task (see Figure 12 panel B).
Figure 12. Automated Operation Span (A), Spatial Delayed Response Task (B-C), Raven’s Advanced Progressive Matrices (D), and Cattell’s Culture Fair Test (E) performance by group before and after training. Error bars represent 95% confidence intervals.
In contrast to these few main effects, none of the cognitive tests administered revealed a time × group interaction effect which would be expected under the hypothesis of differential cognitive test score change by group.

Further analyses with JZS Bayesian repeated measures ANOVAs were largely consistent with the results of these traditional null hypothesis significance tests. Bayes factors comparing the fit of the data under models containing the interaction term (i.e. time × group) versus the model containing only main effects by themselves (i.e. time + group) consistently indicated evidence against the interaction effect for each of the cognitive indices and subtests discussed above. Specific Bayes factors ranged from 1.06 to 9.09, indicating that the observed data are that many times more likely to occur under a model without the interaction effect, versus one that does include it. Bayes factors between 3 and 10 are thought to provide ‘substantial’ (Jeffreys, 1961), or ‘positive’ (Raftery, 1995) evidence against the interaction effect, which describes the pattern of evidence for all but two of the cognitive tests in this case: WAIS-IV Vocabulary subtest (BF$_{01}$ = 1.06), and SDRT spatial maintenance and manipulation (BF$_{01}$ = 2.45). These Bayes factors below 3.0 are thought to offer ‘anecdotal’ (Jeffreys, 1961) or ‘weak’ (Raftery, 1995) evidence against the interaction effect. Further, inspection of the descriptive statistics for these latter two cognitive tests for which evidence against the interaction is weakest, revealed patterns of score change antithetical to gains resulting from training. These include differential decreases in test scores between groups over the training period for WAIS-IV vocabulary, and increases in the no-contact control group scores for the SDRT spatial maintenance and manipulation task. Table 4 lists Bayes factors providing evidence against models including time × group interaction effects for each of the cognitive tests administered.
Table 4.

Bayesian factors for time × group interactions by measure resulting from JZS Bayesian repeated measures ANOVAs

<table>
<thead>
<tr>
<th>Measure</th>
<th>BF₀₁</th>
<th>Quality of evidence against interaction effect</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>WAIS-IV Composite Indices</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Full-Scale IQ</td>
<td>7.51</td>
<td>Substantial</td>
</tr>
<tr>
<td>Verbal Comprehension Index</td>
<td>3.66</td>
<td>Substantial</td>
</tr>
<tr>
<td>Perceptual Reasoning Index</td>
<td>5.22</td>
<td>Substantial</td>
</tr>
<tr>
<td>Working Memory Index</td>
<td>5.90</td>
<td>Substantial</td>
</tr>
<tr>
<td>Processing Speed Index</td>
<td>4.67</td>
<td>Substantial</td>
</tr>
<tr>
<td><strong>WAIS-IV Subtests</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vocabulary</td>
<td>1.06</td>
<td>Weak</td>
</tr>
<tr>
<td>Similarities</td>
<td>8.24</td>
<td>Substantial</td>
</tr>
<tr>
<td>Block Design</td>
<td>4.88</td>
<td>Substantial</td>
</tr>
<tr>
<td>Matrix Reasoning</td>
<td>4.39</td>
<td>Substantial</td>
</tr>
<tr>
<td>Digit Span</td>
<td>3.61</td>
<td>Substantial</td>
</tr>
<tr>
<td>Arithmetic</td>
<td>8.65</td>
<td>Substantial</td>
</tr>
<tr>
<td>Symbol Search</td>
<td>4.56</td>
<td>Substantial</td>
</tr>
<tr>
<td>Coding</td>
<td>6.95</td>
<td>Substantial</td>
</tr>
<tr>
<td><strong>Computerized Working Memory Tasks</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Automated Operation Span</td>
<td>4.16</td>
<td>Substantial</td>
</tr>
<tr>
<td>SDRT Spatial Maintenance Task</td>
<td>7.7</td>
<td>Substantial</td>
</tr>
<tr>
<td>SDRT Spatial Maintenance and Manipulation Task</td>
<td>2.45</td>
<td>Weak</td>
</tr>
<tr>
<td><strong>Nonverbal Reasoning Tasks</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Raven’s Advanced Progressive Matrices</td>
<td>8.51</td>
<td>Substantial</td>
</tr>
<tr>
<td>Cattell’s Culture Fair Test</td>
<td>9.09</td>
<td>Substantial</td>
</tr>
<tr>
<td><strong>Motivation and Expectation</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Motivation to complete training</td>
<td>7.70</td>
<td>Substantial</td>
</tr>
<tr>
<td>Expectation for improvement</td>
<td>6.30</td>
<td>Substantial</td>
</tr>
</tbody>
</table>

Note: SDRT, Spatial Delayed Response Task
Thus, these results suggest that while participants showed facilitation of performance at the second administration after training on some cognitive tasks, none of these effects were observed to significantly vary by group. Figures 11 and 12 display these pre- versus post-training cognitive scores.

3.3.1.4 Motivation to Train and Expectations for Improvement

Analysis of participants’ self-reported motivation to complete online training, as well as the degree to which they thought they might improve on the training tasks over the course of the training period did not reveal any significant time × group interactions. Results of the mixed-design repeated measures ANOVAs indicated main effects of time for both motivation to complete training \[F(2,84) = 19.40, p < .001\], and expectations for improvement \[F(2,84) = 5.83, p = .004\]. Bayesian analyses were carried out on these measures as well, and indicated strong evidence against the interaction effect of time × group: \( BF_{01} = 7.70 \) for motivation to complete training, and \( BF_{01} = 6.30 \) for expectation for improvement. Thus, participants in both the WM training group and PS active control group indicated a decline in motivation across the training period, but not at significantly different rates. Self-reported ratings of expectations for improvement followed a U-shaped curve for both groups, with lowest expectations for improvement mid-way through training. Figure 13 displays these metrics across the training period.
3.3.1.5 Training Time Correlations

Interestingly, despite overall non-significant findings concerning time × group interaction effects for cognitive test scores, correlation analysis of total time spent training reveals differences between groups, and potential individual differences within the WM training group. Specifically, the total amount of time members of the WM training group spent training was significantly correlated with gains in measured WAIS-IV FSIQ \((r = .42, p = .039)\), however, not for any of the constituent composite indices (VCI, PRI, WMI, PSI; \(r’s = .13 - .38, p’s = .06 - .28\)), nor intermediate measures of working memory ability (AOSPAN task and SDRT; \(r’s = -.31 - .31, p’s = .14 - .89\)), nor either measure of nonverbal ability administered (RAPM, CCFT; \(r’s = -.07 - .06, p’s = .75 - .76\)). Conversely, total time spent training by members of the PS active
control group was not found to be significantly associated with observed gains in FSIQ, nor any of the above listed measures ($r' = -.36 - .33, p' = .08 - .81$) with the one exception of VCI ($r = .42, p = .039$).

### 3.4 Discussion

The goal of the present study was to evaluate the weight of evidence for or against the controversial claim that WM training ‘works’; or more specifically that training of WM transfers to untrained cognitive tasks in the domain of fluid intelligence. We evaluated this hypothesis in a community-recruited sample of healthy young adults, aged 18-40, in a single-blind randomized controlled six week trial of online WM training compared to both active and no-contact control groups. Experimenters and assessors were blinded to participant group assignment to the largest extent possible, however several logistical indicators regarding individual participation (e.g. fewer appointments for no-contact controls) unfortunately prevented a truly double-blind experimental design.

The present results provide no convincing evidence of near-transfer of working memory training to working memory capacity, or far-transfer to fluid intelligence. While participants in both the WM training group and the PS active control group were observed to improve their performance substantially on the Lumosity training tasks themselves, this facilitated performance did not generalize, or demonstrate near-transfer to the closely related domains of working memory capacity, or to visual working memory. Similarly, improved performance on the WM training tasks did not demonstrate far-transfer to measures of fluid intelligence, nor to a broad range of cognitive domains measured by a traditional comprehensive test of intelligence. Stated plainly, participants randomized to six weeks of online working memory training fared no better on these cognitive tasks after training, when compared to those randomized to a processing speed
active control condition, or even compared to those randomized to a no-contact control condition. Several cognitive tests and indices evinced higher scores at the post-training cognitive assessment relative to the pre-training assessment (e.g. WAIS-IV PRI, PSI; AOSPAN; RAPM; CCFT); however, in each case, the effect did not significantly differ by group, suggesting practice effects for the tests themselves versus true training-related gains in performance (see Salthouse & Tucker-Drob, 2008). Overall, this pattern of results supports our first hypothesis (that participants would improve on training tasks), though provides substantial evidence against our more consequential second hypothesis (that WM training would precipitate near-transfer to WM capacity), and third hypothesis (that WM training would precipitate far-transfer to fluid intelligence).

Counter to these results, post-hoc analyses revealed that total time spent training by members of the WM training group was positively and significantly correlated in observed gains in overall intelligence as measured by the WAIS-IV full-scale intelligence quotient (FSIQ) index. This pattern did not obtain for the PS active control condition. However, two indicators suggest that this finding should be interpreted with caution, if not completely disregarded. First, similar correlations did not hold for constituent indices of the FSIQ (i.e. VCI, PRI, WMI, or PSI). Second, total time spent training by members of the PS active control group was positively and significantly correlated with gains on WAIS-IV VCI (composed of tests of vocabulary and abstract verbal reasoning) for which there is no theoretical basis for improvement following training of processing speed. Rather, both of these correlations are more than likely spurious, resulting from measurement error and/or psychometric imprecision (discussed below).

Looking to the literature, these results are consistent with a large and growing body of empirical work in support of the null for WM training (Colom et al., 2010; Owen et al., 2010;
Chooi & Thompson 2012; Redick et al., 2013; Sprenger et al., 2013; Thompson et al., 2013; Harrison et al., 2013; Colom et al., 2013; von Bastian & Eschen, 2016; Lawlor-Savage & Goghari, 2016). However, due to the largely divided or ‘reliably ambiguous’ (Urbánek and Marček; 2016) nature of the current WM training literature, these results are also inconsistent with a large and growing opposing body of empirical work that has demonstrated evidence for both near- and far-transfer effects resulting from WM training in healthy adults (Jaeggi and colleagues, 2008, 2010, 2011; Chein & Morrison, 2010; Schmiedek et al., 2010; Schweizer et al., 2013; von Bastian & Oberauer, 2013; Buschkuehl et al., 2014; Hardy et al., 2015).

While the present results land firmly and unambiguously on the former side of this split literature, the addition of our single empirical result cannot hope to ultimately settle the debate on WM training efficacy. However, a more targeted comparison of study methodology may provide several clues as to why it found support for the null. For example, following Melby-Lervåg and colleagues (2016) analysis, narrowing the broader WM training literature to only the 34 comparisons to date which have included 20 or greater participants per group, and also utilized an active control condition revealed a negligible mean effect size. In comparison, every other combination of experimental design (e.g. < 20 participants per group, with untreated controls etc.) yielded significant mean effect sizes. In other words, the literature composed of methodologically rigorous studies is not so split or divided as the broader WM training literature, and the present results are indeed consistent with these similarly rigorous experimental trials.

Despite methodological rigor on these important points, limitations of the current study include equivalence of pre- and post-training cognitive test forms, as well as a high degree of participant attrition from the both the WM training group and the PS active control group. First, regarding the equivalence of test forms, here we split singular tests into roughly equivalent
versions according to even and odd item numbers. However, because most of these cognitive tests are designed such that each successive question is incrementally more difficult than the last, it remains possible that the form containing even-numbered items is slightly more difficult than the one containing odd-numbered items despite good psychometric properties in terms of split-half reliabilities. In the present experimental design, we decided on the most conservative option which is to use the odd-numbered items at pre-training assessment, and even-numbered items at post-training assessments. Due to our relatively small sample sizes of ~25 per group, and the possibility for participant attrition, we decided not to randomize or counterbalance test forms across pre- and post-training assessments. While it is possible that administering the potentially more difficult tests at post-training assessment served to dampen any training-induced gains in cognitive ability, this possibility is not supported by the equivalence of the no-contact control group, or the absence of time × group interaction effects on tasks which were not split into alternate forms but rather administered in their entirety before and after training (i.e. WAIS-IV Digit Span, AOSPAN, SDRT).

Regarding participant attrition, it should be noted here that while only 7, and 8 participants withdrew from the Lumosity WM training and PS active control conditions (or abandoned their prescribed training plan) after randomization respectively, these numbers represent a rather large proportion of the total group sizes (7/32 = 21.89% for the WM group, and 8/32 = 25% for the PS group). This drop-out may speak to any number of factors about the tolerability of the Lumosity interventions, and leaves the current results open to speculation about potential systematic differences between trial completers and trial abandoners. Anecdotally, several participants noted in conversation with the experimenters that training became less exciting and somewhat repetitive across the six week training period. These
sentiments are corroborated quantitatively for both the WM training group and PS active control
group by substantial decreases in self-rated motivation and expectations of improvement from
training between the start of the trial and even halfway through. Several participants (from both
the WM and PS groups) expressed that adding more variety to the training regimen may have
served to enhance its appeal. Regardless of whether the repetitive nature of the highly
circumscribed sets of Lumosity training tasks accounts for any of the participant drop-out,
Straus, Glasziou, Richardson & Haynes (2011) discuss the implications of attrition from
randomized controlled trials, and point out that many medical journals will refuse to publish
trials with attrition rates above 20%. Examination of the factors that lead to WM training
adherence and attrition will be important topics of future research (see Double & Birney, 2016).

Finally, while the current study includes just over the minimum number of 20 participants
per group recommended by the literature (Simmons et al., 2011), it should be pointed out that
power analyses are highly sensitive to estimates of effect size, and particularly so at the lower
end. For example, based on an early estimate of a small to moderate mean effect size of \( d = 0.34 \)
\( (f = 0.17) \) for n-back training (Melby-Lervåg & Hulme, 2013), experimental designs would
require 12 participants per group in order to achieve a power of 0.8 with \( \alpha = .05 \) in a two group ×
two time repeated measures ANOVA design with a test re-test reliability of 0.85 for the measure
in question (Faul, Erdfelder, Buchner, & Lang, 2007; Faul, Erdfelder, Buchner, & Lang, 2009).
However, if the estimated effect size drops to just \( f = 0.1 \), that same experimental design would
require 31 participants per group to maintain an 80% chance of rejecting the null hypothesis if it
were indeed false. The danger here of course is that it is the true effect size of WM training itself
that is in question in the literature. Basing study design on artificially high effect size values
could have the effect of lowering true values for statistical power, leading to poorer chances of
detecting effects if they truly exist, and also poorer chances that any found effects are indeed genuine (Button et al., 2013; Bogg & Lasecki, 2015; Moreau, Kirk, and Waldie, 2016). Future research on WM training efficacy will benefit from greater statistical power resulting from larger sample sizes, and more accurate estimates of effect size. Online tools for homogenizing study design, and streamlining participant training will certainly aid in organizing larger multi-site WM training studies (see von Bastian, Locher & Ruflin, 2012 for an early example).

These limitations notwithstanding, our trial includes a number of strengths that work to improve upon methodological shortcomings that have been described as ubiquitous or pervasive in the existing WM training literature (Button et al, 2013; Boot, Simons, Stothart, & Stutts, 2013; Redick, 2015; Melby-Lervåg et al., 2016; Moreau et al., 2016). In addition to utilizing minimum suggested sample sizes, and employing an active control condition, the present study sought to reduce the ambiguity of potential positive findings by measuring a number of intra-individual variables that have been suggested to moderate WM training effect, including: self-rated motivation to complete training, self-rated expectations of cognitive improvement from training, major personality traits, grit, need for cognition, as well as current cognitive activities. By measuring and ensuring equivalence between groups on these potentially important intra-personal variables, in addition to vital demographic characteristics (i.e. age, sex, education, and IQ), their impact on any potential gains in cognitive ability can be effectively ruled out. No such gains in ability were observed in this case, however because these traits were measured, we can state with some confidence that our null findings were not due to unmeasured differences in these variables between our three groups. The near-perfect equivalence of our three groups on all of the above variables precludes the necessity to statistically model pre-training group differences in our analyses. Additionally, and contrary to much of the previous literature, we
utilized multiple measures of the cognitive domains of interest: working memory (WAIS-IV Digit Span, and Arithmetic, AOSPAN, SDRT), and fluid intelligence (WAIS-IV Matrix Reasoning, and Block Design, RAPM, and CCFT which is composed of four separate tests of Gf ability). Importantly, each of these measures within these given domains of interest returned consistent results in support of the null regarding WM training.

A final strength of our trial is that cognitive test scores were not observed to decrease over the course of the training period for either of the control groups, which Redick (2015) has pointed out as a contributing factor to significant time × group interactions in several successful WM training studies. It is interesting to point out here however, that while including an active control condition that closely matches all but the proposed intervention of the treatment group is certainly a methodological asset (Boot et al., 2013), our active and passive control groups obtained very much the same result – i.e. no significant improvement on any cognitive test which could not otherwise be due to expected practice effects. This is an interesting and somewhat unexpected result given the large discrepancies in average mean effect sizes for trials with treated versus untreated controls listed in meta-analytic reviews. Recall that Melby-Lervåg and colleagues (2016) found effect sizes of 0.15 and 0.26 for n-back training on nonverbal ability for treated and untreated controls, whereas Au and colleagues (2014) found an even larger discrepancy with effect sizes of 0.06 and 0.44 for treated versus untreated controls in their more targeted review. Heterogeneity of study design in the WM training literature makes it difficult to compare the equivalence of our active and passive control conditions to previous studies. An in depth examination of Melby-Lervåg and colleagues (2016) supplementary material yielded no comparable studies meeting the following criteria: 1) sample of young adults (vs. children or older adults); 2) 20 or greater participants per group; 3) participants randomized to both active
and passive control groups in addition to the treatment group(s); 4) utilization of the dual n-back task for training; and 5) examination of fluid intelligence as an outcome measure. The closest experimental trial to these criteria is that of Redick et al. (2013), which meets all of the above conditions except true group randomization. Interestingly, their results indicated a similar pattern to those found here: non-significant differences between all three groups, including both active and passive control conditions. These results raise the thorny question of whether other trial- or researcher-specific factors may account for some of the variance observed between studies which include active control conditions, and those that do not (e.g. experimenter bias, publication bias etc.). Notably, Redick et al.’s (2013) trial also shares in common with the current study, the failure to find near-transfer of training to measures of WM span, or WM capacity, contrary to many findings to this effect in the literature (Jaeggi et al., 2010; Morrison & Chein, 2011; Shipstead, et al., 2012).

In sum, the present study found no convincing evidence of far-transfer of WM training to untrained measures of $Gf$, nor near-transfer of training to intermediate cognitive domains (i.e. WM capacity) thought to mediate increases of $Gf$ in young adults. Importantly, we implemented a methodologically rigorous design following recommendations from recent literature, and also measured a variety of intra-personal factors that have been proposed to moderate treatment effect. Overall, while the present results in support of the null hypothesis cannot hope to singly resolve the heated debate over the controversial claims of WM training efficacy, they do contribute meaningfully to the rapidly growing corpus of research on the topic. Crucially, by providing additional and incremental evidence against the efficacy of dual n-back training in healthy young adults, subsequent research can intensify the search for alternative interventions that may produce the desired effects in this population (see McCabe, Redick, & Engle, 2016), or
alternative populations or patient groups for which dual n-back training may actually be effective (see Ansari, 2015 for a review, and Weicker et al., 2016 for a meta-analysis).
3.5 References


doi:10.1177/0013164490504019


doi:10.1126/science.1736359


Chuderski, A. (2013). When are fluid intelligence and working memory isomorphic and when are they not? *Intelligence, 41*(4), 244-262. doi:10.1016/j.intell.2013.04.003


111


JASP Team. (2016). JASP (version 0.8.0.0) [computer software]. Retrieved from https://jasp-stats.org/


Chapter Four: Neural Mechanisms of Working Memory Training and Transfer

4.1 Background

Successful completion of a wide variety of complex tasks such as reasoning and problem-solving rely to a large degree on the underlying cognitive process of ‘working memory’ (WM), or the ability to maintain and mentally manipulate task-relevant information over short periods of time (Baddeley, 1992, Repovs & Baddeley, 2006). Individuals vary in terms of how many pieces of information they can hold in WM (i.e. WM capacity; Engle, Kane, & Tuholski, 1999), and these individual differences have in turn been variously associated with performance on a wide variety of academic and occupational outcomes in addition to reasoning and problem-solving, including: reading comprehension, language and mathematic abilities, as well as overall academic performance, and even ‘fluid intelligence’ more broadly (Engle, Tuholski, Laughlin, & Conway, 1999; Eriksson et al., 2015). Fluid intelligence (denoted Gf) has been conceptualized as the ability to adapt one’s reasoning abilities to solve novel cognitive problems, and stands in contrast to ‘crystallized intelligence’ which conversely draws heavily on previously learned declarative information (Cattell, 1963; Carroll, 1996; McGrew, 2009).

Although WM capacity has traditionally been thought of as a stable trait which is cemented before early adulthood, recent research has suggested that training of WM specifically may lead to gains in WM capacity, which are then thought to further generalize or transfer to related important domains such as Gf (see Jaeggi, Buschkuehl, Jonides & Perrig, 2008; Jaeggi, Studer-Luethi, Buschkuehl, Su, Jonides & Perrig, 2010; Jaeggi, Buschkuehl, Jonides & Shah, 2011 for early examples). However, these early examples of WM training and transfer success were quickly met with empirical counterexamples of WM training failure, or in some cases, success in terms of near-transfer of training to WM capacity or the training tasks themselves, but
failure of far-transfer to untrained tasks in more distal domains of cognitive functioning such as Gf (see Redick et al., 2012, 2013; Thompson et al., 2013). Despite (or perhaps due to) this disagreement about the basic effect in question, and the potential therapeutic effects of essentially increasing task-general cognitive abilities, the WM training literature has rapidly expanded in recent years to include multiple reviews (e.g. Buschkuehl & Jaeggi, 2010; Conway & Getz, 2010; Takeuchi, Taki, & Kawashima, 2010; Morrison & Chein, 2011; Shipstead, Redick, & Engle, 2012; von Bastian & Oberauer, 2013, Redick, Shipstead, Wiemers, Melby-Lervåg, & Hulme, 2015), and meta-analyses (Melby-Lervåg & Hulme, 2012; Au, Sheehan, Tsai, Duncan, Buschkuehl, & Jaeggi, 2014; Schwaighofer, Fischer, & Buhner, 2015; Weicker, Villringer, & Thöne-Otto, 2016). Unfortunately however, these meta-analyses have been inconsistent in terms of methodology and study inclusion, leading the authors to differing conclusions regarding the basic questions of whether or not WM training ‘works’ – and if so, for whom?

In light of this ‘reliably ambiguous’ literature (Urbánek and Marček, 2016), a parallel approach to the study of WM training efficacy has been to investigate the neural correlates of WM task performance before and after such training, in an effort to discern whether or not training has any effect at the neural level. Indeed, a working assumption of the literature has been that for WM training to be effective, tasks used for the training itself must share at least some components with the outcome tasks, in terms of cognitive capacity constraints (see Halford, Cowan, & Andrews, 2007), or underlying task-based neural circuitry (Buschkuehl, Jaeggi, & Jonides, 2012, Dahlin, Neely, Larsson, Bäckman, & Nyberg, 2008; Dahlin, Bäckman, Neely, & Nyberg, 2009). Thus, in addition to adding insight to the potential behavioural effects of WM training, investigating associated neural mechanisms may help in identifying elements of
successful trials, developing more specific targets for intervention (Brugess, Gray, Conway, & Braver, 2011), and also in generally offering a better understanding the functional brain plasticity that might support such behavioural improvement in higher cognitive processes (Thompson et al., 2016).

Indeed, very little is known about the correspondence between the expansion of WM capacity, and the neural systems that are known to support WM processes more generally. Neuroimaging studies have reliably demonstrated that activation in dorsolateral prefrontal cortex and posterior parietal cortex scales with the amount of information required to be held in memory (Braver et al., 1997), and further that individual differences in WM capacity are correlated with load-dependent activations in these areas (McNab & Klingberg, 2008). Meta-analyses of previous neuroimaging studies have revealed activation of a circumscribed fronto-parietal network in response to WM task demands through proposed mechanisms of persistent activation (Curtis & D’Esposito, 2003), including: 1) bilateral dorsolateral prefrontal cortex; 2) bilateral mid-ventrolateral prefrontal cortex; 3) bilateral rostral prefrontal cortex; 4) bilateral premotor cortex; and, 5) bilateral and medial posterior parietal cortex (reviewed in Owen, McMillan, Laird, & Bullmore, 2005; Rottschy et al., 2012). Interestingly, similar frontal and parietal areas of activation in response to task load have been noted in measures which WM training has been purported to transfer to (i.e. measures of Gf; see Clark, Lawlor-Savage, & Goghari, in preparation for a review). Additionally, and consistent with the strong behavioural associations between WM, Gf, and general intelligence, these fronto-parietal areas of activation are highly consistent with Jung & Haier’s (2007) Parieto-Frontal Integration Theory (P-FIT) of intelligence, which attempts to neuroanatomically localize general cognitive ability. Briefly, the P-FIT model suggests that the majority of information processing (i.e. that which differentiates
persons of high and low intellectual ability) occurs in a complex interplay between frontal and parietal cortical areas, while basic visual and auditory information is collected by the temporal and occipital lobes. Jaušovec & Jaušovec (2012) summarized a small literature focusing on activation differences between people of high and low intelligence under cognitive load. Notably, highly intelligent people appear to make relatively greater use of parietal areas compared to frontal areas, whereas people of lower intelligence showed the opposite pattern. This finding is consistent with earlier reports of higher ability people using a left hemisphere sub-network, including the inferior parietal lobule and Broca’s area, in an n-back WM training paradigm, compared to low ability people who made use of a right-hemisphere sub-network including the inferior parietal lobule and connections to the dorsolateral prefrontal cortex (Glabus et al., 2003). Furthermore, Jaušovec and Jaušovec (2012) found that highly intelligent people also appear to show greater activation in the earlier stages of task completion compared to those of lower intelligence who displayed the opposite pattern. Finally, and perhaps expectably, on tests of graded difficulty high-ability people tend to show an increase in activation progressing from easy to more difficult tasks, whereas low-ability people showed the opposite pattern (also see Jaeggi, Buschkuehl, Etienne, Ozdoba, Perrig, & Nikko, 2007). Overall, this body of work implicates individual differences in fronto-parietal network structure and function in general cognitive ability and associated cognitive domains, such as WM and Gf.

However, despite the approximate neuroanatomical synchrony of basic WM, Gf and general intellectual processes, very little is known about how activation of this circumscribed fronto-parietal network might be expected to change following WM training. Buschkuehl, Jaeggi, & Jonides (2012) provide a carefully considered theoretical framework for all the possibilities. The first possibility is activation increase in these WM task-related cortical areas.
Increases in brain activation are common following training in sensory and motor tasks, and are taken to represent stronger neural responses in associated areas, or to increased size of cortical representations. Buschkuehl et al. (2012) note that activation increases following WM training specifically are quite rare. The second possibility is activation decrease in these same WM task-related areas, which has been associated with increased efficiency of neural circuits involved in a trained task – or, the non-use of brain areas irrelevant to task performance. Indeed, in a review of studies on neural efficiency, Neubauer & Fink (2009) note an overall negative correlation between overall intelligence and brain activity under cognitive load. A third possibility represents a combination of the first two – both increases and decreases in activation across different brain regions, or activation redistribution. Buschkuehl et al. (2012) note that while activation decreases in the anterior cingulate cortex might be explained in terms of the reduced requirement for attentional control following training, activation decreases in other areas are somewhat more difficult to explain. A final way in which brain activation may respond to cognitive training is recruitment of entirely new areas of the brain altogether, or activation reorganization. While a small corpus of studies have found such patterns following various cognitive training regimens, Buschkuehl et al. (2012) note that such a pattern has not been reported following WM training specifically.

Critically, Kelly and Garavan (2005) discuss the above possibilities for neural change in the context of practice effects for given cognitive tasks. Specifically, they suggest that completion of novel tasks is associated with control and attention areas necessary to focus on the novel task. These include the prefrontal cortex, the anterior cingulate cortex, and the posterior parietal cortex. With practice, less attentional control is needed in order to complete a task, and these areas accordingly become less active during task completion as a function of practice.
Kelly and Garavan (2005) use this insight to propose a crucial distinguishing criterion between activation redistribution, and true activation reorganization:

If the cognitive processes are changed by practice such that practised task performance amounts to performance of a cognitively different task, this is likely to be reflected in a neurobiologically different task map and therefore a true reorganization of functional activations. On the other hand, if the cognitive operations underlying task performance are not fundamentally changed by practice, but rather the amount of control and attention support is altered, then the pattern of activation changes observed is likely to be that of redistribution. (pp.1091).

Given this overview of the theoretical possibilities for activation changes in response to cognitive training, we can turn our attention to the very small set of studies which have examined relative changes in activation of this network in response to WM training specifically, in samples of healthy young adults. An early fMRI investigation of WM training by Olesen, Westerberg, & Klingberg (2004) found increases in activation in frontal and parietal cortices in participants following five weeks of WM training in small samples of young adults (experiment 1, N = 3; experiment 2, N = 8). Specifically, the right middle frontal gyrus, right inferior parietal cortex, and bilateral intraparietal cortices showed an increase in task related activation following training. However, the researchers also reported specific areas of decreased activation following training, including the cingulate sulcus, the inferior frontal sulcus, and the postcentral gyrus. While transfer or generalizability of WM gains to other cognitive domains was not investigated, the results suggested that changes on the neural level are indeed happening over the course of WM training, and in the general frontal and parietal regions that would be expected. In contrast to these increases in fronto-parietal activation, subsequent investigations have generally revealed the opposite pattern, or decreased activation in fronto-parietal networks following some form of WM training. For example, Schneiders, Opitz, Krick, & Mecklinger (2011) trained a sample of
young adults ($N = 48$) with either and auditory or visual adaptive n-back for 8-10 ~1-hour training sessions over a two week period. Pre- and post-training fMRI data collected on visual 0-back and 2-back tasks revealed significant activation decreases in the right superior middle frontal gyrus, as well as right posterior parietal lobule regardless of training condition. Further decreases in activation were observed in the right middle frontal gyrus for the visual n-back training group only. Schneiders and colleagues attributed these activation decreases to increased neural efficiency of intra-modal processes (i.e. visual), as well as across-modal processes (i.e. improved general control from either type of training). In a follow-up study, Schneiders and colleagues (2012), trained a separate sample of young adults ($N = 32$) with an adaptive auditory n-back task across eight ~50 minute training sessions across two weeks. Again, pre- post-training fMRI data was collected during performance of both visual and auditory 0-back and 2-back tasks. Consistent with their previous results, participants showed decreased activation in inferior parietal regions, and the right superior frontal gyrus during both auditory and visual n-back task performance. However, additional decreases in activation were observed in the right inferior frontal gyrus during auditory n-back performance only, providing additional evidence of greater activation decreases in response to intra-modal training.

More recently, Schweizer, Grahn, Hampshire, Mobbs, & Dalgleish (2013) trained young adult participants ($N = 24$) on an emotionally salient variant of the adaptive n-back task for 20 sessions lasting ~20-30 minutes each. Pre- and post-training fMRI data revealed activation decreases in the left dorsolateral prefrontal cortex, right superior frontal gyrus, bilateral supramarginal gyrus, and bilateral middle occipital lobe for the 3-back condition. Interestingly however, they found only activation increases when examining the 5-back condition, which they attributed to increased effort in the 5-back condition at post-training, which is consistent with
previous investigations of WM network activation in response to increased WM task load. Results from Buschkuehl, Hernandez-Garcia, Jaeggi, Bernard, & Jonides (2014) provide further support for the notion of activation increases in response to difficult tasks, and activation decreases in response to easier, or more adequately trained tasks that require less executive or attentional control. Specifically, following training of a sample of young adults ($N = 55$) for 20 minutes per day for seven consecutive days on an adaptive visuospatial n-back task, Buschkuehl and colleagues (2014) found greater activation in response to the 4-back condition compared to before training in the right frontal cortex (Broadmann area 6), and occipital cortex (Broadmann area 19). The authors explain these findings in the context of the dual-process theory of human performance (Chein & Schneider, 2005; Posner & Snyder, 1975) which posits that initial training on a new task requires a greater amount of effort and resources compared to one which has been trained, and become more automatic in execution. For this reason, Buschkuehl and colleagues speculated that a longer duration of training in their sample would have led to the more familiar pattern of fronto-parietal network activation decreases at post-training neuroimaging. In line with this prediction, one of the earliest examples of an fMRI investigation of WM training revealed a complex time course for training related changes in neural activation for a small sample of healthy young adults ($N = 9$) by scanning participants midway through their training in addition to before and after. Hempel and colleagues (2004) reported an inverse U-shaped quadratic function for cortical activation across four weeks of twice-daily (specific duration not reported) visuospatial adaptive n-back WM training. Task performance increased linearly across training, but was associated with both activation increases in the inferior frontal gyrus, and intraparietal sulcus just two weeks into training, and decreases in these same areas at the end of four weeks of
training. The authors suggested that the prefrontal and parietal cortices may facilitate two parallel mechanisms of WM, one which enhances active WM, and another that supresses past stimuli.

Most recently, Thompson et al. (2016) examined changes in neural functioning in response to 20 days of WM training in a sample of healthy young adults ($N = 58$) in a randomized controlled design which utilized both treated and untreated control groups as comparisons. The experimental group trained with a typical adaptive dual n-back task (i.e. similar to that employed by Jaeggi et al., 2008), and the active control group practiced with a visuospatial training task involving multiple object tracking. Behavioural results were in line with previous literature indicating increased n-back task proficiency for the n-back training group, but not for either the treated nor untreated control group. Also in line with previous literature, neuroimaging results showed reduced neural activity in response to 2-back and 3-back versions of the n-back task in fronto-parietal networks after training for the n-back training group only. Importantly, because Thompson and colleagues (2016) included and scanned an active control condition, they were the first to show convincingly that the decreases in activation in response to WM training were not simply due to the effect of any intensive or adaptive training program. However, despite the expected patterns of behavioural performance increase on the n-back task itself, and the accompanying neural activation decreases in key cortical areas that support WM processes, WM trained participants did not evince far-transfer to any other domain of cognition (reported in Thompson et al., 2013).

Given this review of the literature, the logic behind further investigating the neural mechanisms of WM training and transfer to untrained domains of cognition becomes clear. Beyond simply attempting to replicate the positive behavioural far-transfer results of Jaeggi et al. (2008) and others, combined behavioural and neuroimaging techniques offer the prospect of
gauging the neural changes precipitated by WM training, not only for the training task itself, but perhaps more importantly, for the transfer task(s). That is, while previous investigations have shown adaptation of the fronto-parietal network resulting from WM training, to our knowledge, no study has yet examined whether similar changes in activation occur in the *untrained* tasks to which that WM training is purported to transfer (such as Raven’s Matrices, described below). Thus, in addition to adding insight to a basic understanding of how fronto-parietal networks crucial in general intellectual functioning react to WM training, the present trial also seeks to approach the central question of WM training efficacy from a neuroscience perspective. Stated simply, if WM training does indeed transfer to ability in untrained cognitive domains, we should see evidence of this far-transfer in terms of altered activation patterns in response to untrained tasks following training.

Regarding this primary research question of interest, and following the results of Thompson et al. (2016), we predict: 1) reduced fronto-parietal activation for difficult levels of the dual n-back training task (described below) following WM training, but not processing speed (PS) training, which we utilized here as an active control condition. Both behavioural and neuroimaging investigations support PS training as a viable active control for WM training, by holding constant the level of effort, motivation, interaction with computers and researchers, while impacting relatively orthogonal behavioural skillsets and neural functions (Takeuchi et al., 2011; Takeuchi & Kawashima, 2012). Additionally, following the logic that both WM and *Gf* tasks rely heavily on the fronto-parietal network (see Clark et al., in preparation¹), we predict: 2) post-training reductions in fronto-parietal activation for difficult levels of our transfer task, Raven’s Standard Progressive Matrices for the WM training group, but not the PS active control group. Finally, 3) we expect to find no changes following WM or PS training on a third scanned
task of lexical decision making, which theoretically ought not to be facilitated by either type of training. The lexical decision task was included as a comparison task in order to assess the specificity of training-related changes in activation in the domains of WM and Gf.

4.2 Method

4.2.1 Participants and Recruitment

A total of 359 healthy adults were recruited via advertisements and physical postings distributed throughout the community. Participants were directed to complete online screening measures at braintrainingstudy.ca which inquired about the following exclusion criteria, including: 1) age less than 18, or greater than 40; 2) left-handedness; 3) history of traumatic brain injury or other neurological condition causing sensory or motor impairment; 4) self-reported of presence Axis I mental illness; 5) less than normal or corrected-to-normal visual acuity; and 6) MRI contra-indications; 7) insufficient access to a computer and high-speed internet; and 8) recent or previous use of the n-back training task or other online cognitive training paradigms. A total of 187 participants were invited to participate in the study out of the 359 participants who completed the screening questionnaires. Of these, 49 participants were included in the final neuroimaging analyses. See Figure 14 for a flow chart illustrating participant recruitment, randomization, and exclusion.
Figure 14. Flow chart of study design. *Two participants in MRI conditions were reassigned to the no-contact control group after being unable to tolerate MRI scanning. †Participants removed from analysis due to training contamination, low training dosage, or data acquisition issues.
Participants received compensation of $20 per cognitive testing session, and $20 per MRI session, for a total of $80 across the four required appointments (pre- and post-training cognitive testing and MRI sessions). All participants provided written consent to participate in the study, and ethics approval was obtained from the University of Calgary’s Conjoint Health Research Ethics Board (CHREB).

4.2.2 Procedure and Materials

Once recruited, screened and invited for participation in the study, participants were randomized to a working memory training group ($n = 25$), or a processing speed active control group ($n = 24$). A no-contact control group was also recruited ($n = 27$) for comparison of behavioural data, but did not participate in the neuroimaging component of the trial (details of behavioural analyses pertaining to this no-contact control group can be found in Clark, Lawlor-Savage, & Goghari, in preparation). Participants were blinded to group randomization, however, experimenters were in several cases made aware of group randomization when participants disclosed particular training games pertaining to their assigned training condition. For this reason, unfortunately the trial does not conform to the highest standards of double blind experimental procedures. After participants were assigned to groups, members of the WM and PS training group first underwent their initial MRI session, followed soon thereafter by their initial cognitive testing appointment on a separate day. Following these appointments, they were given access to pre-made research accounts on Lumosity.com (Lumos Labs Inc., 2009), which limited access to the prescribed WM or PS training games (described below). Participants were asked to engage with their respective training programs at home for 20 minutes of per day, for five out of seven days per week, for six weeks. Progression through training was monitored via reports from Lumosity which contained logs of participant engagement across all games in their
training program. Individuals were removed from the trial if they 1) did not complete at least 20 out of the total 30 assigned days of training, or; 2) if they had contaminated their training program by accessing games other than those specifically selected for WM or PS training. Participants in these training groups were encouraged and reminded to continue with their training regimen via weekly emails sent by the experimenters. After the six week training period, participants in the WM training and PS active control groups underwent a second, post-training, MRI scanning session and cognitive testing appointment shortly thereafter on a separate day. While the no-contact control group participated in cognitive testing, the present investigation focuses on the neuroimaging results for the WM training and PS active control condition exclusively. Comparison of behavioural cognitive testing results for all three groups is reported and discussed in Clark et al. (in preparation²).

4.2.2.1 Cognitive Testing and Behavioural Measures

Pre- and post- training cognitive testing was completed as part of a larger WM training behavioural trial, and included a number of cognitive measures not reported or discussed here. Further discussion of these measures and analyses regarding pre- to post-training changes can be found in Clark et al. (in preparation²). Here, we report only pre-training split-half full scale intelligence quotient (FSIQ) values from the Wechsler Adult Intelligence Scale – fourth edition (WAIS-IV) to demonstrate cognitive/intellectual equivalency of our groups before training. All cognitive assessments were completed by PhD-level graduate students with specific training in neuropsychological assessment, or trained undergraduate volunteers. In addition to the above cognitive assessments, participants were also asked to complete questionnaires on a wide variety of other variables which might mediate or moderate observed effects of online cognitive training: including: measures of personality (HEXACO; Ashton & Lee, 2009), need for cognition
(Cacioppo, Petty, & Kao, 1984), ‘Grit’ (i.e. commitment to long term goals; Duckworth & Quinn, 2009), and current cognitive activities (Eskes, et al., 2010). Participants in the WM training and PS active control groups were also asked to complete training-specific measures of motivation to complete training, and expectations of cognitive improvement as a result of training. Measuring motivation as well as the expected cognitive benefits of training is particularly important given the literature regarding the potential for motivational factors to artificially facilitate training effects (see Boot et al., 2013; Foroughi et al., 2016). All questionnaires were administered once at the beginning of the study, except the motivation and expectancy questionnaire which was administered before, mid-way through, and after training.

4.2.2.2 Neuroimaging Tasks

MRI sessions included structural scans for anatomical registration, as well as functional scans of three separate cognitive tasks: 1) dual n-back task; 2) Raven’s Standard Progressive Matrices; and 3) a lexical decision task. Participants in the WM training and PS active control groups were introduced to these tasks before each MRI session, and were asked to practice with them until they indicated full comprehension of task instructions and requirements. Practice trials included immediate feedback for responses, whereas the in-scanner trials did not. While in the scanner, participants viewed stimuli on a screen suspended behind their heads in the magnet bore via a mirror attached to the head coil. Participants were additionally fitted with headphones for presentation of audio stimuli, and utilized a scanner-compatible two-button response box placed in their right hand in order to indicate their responses to task trials with their index and/or middle fingers. All stimuli were presented electronically using E-Prime software (Psychology Software Tools, 2012). MRI sessions typically ranged from 60 to 75 minutes in duration. The three tasks
described below were completed in a random order for each participant, though task order was kept constant between pre- and post-training MRI sessions.

4.2.2.2.1 Dual n-Back Functional Imaging Task

Jaeggi and colleagues’ (2008) description of the dual n-back training task was followed as closely as possible in the current implementation, though was modified for use in the scanner. For each trial in a block-design, participants viewed a blue square presented at one of eight locations around the periphery of a 3x3 grid with a fixation cross at centre, and simultaneously listened to an auditorily presented letter (D, F, H, J, M, Q, R, or X). Visual stimuli were presented for 500 milliseconds (ms), followed by 2500 ms of fixation which also served as the response window for that trial. Participants indicated a visual match to n trials previous with their index finger, auditory match to n trials previous with their middle finger, and combined visual and auditory match with both fingers. No response was required on trials not containing a match. Participants underwent four blocks each of 1-back, 2-back, and 3-back conditions, where each block included 20+n trials. See Figure 15, panel A.
Figure 15. In-scanner cognitive tasks: 2-back condition of the n-back task (A); adapted Raven’s Standard Progressive Matrices (B); and, the lexical decision task (C), with correct answers indicated.
Visual stimuli location, and letter presentation were randomized within each block, though each block was structured such that it contained four visual matches, four auditory matches, and two combined visual plus auditory matches. Each block was preceded by 5000 ms of instructions, and separated from subsequent blocks by 15 seconds of fixation in order to allow the hemodynamic response to return to baseline. Block order was fixed for all participants: 1-back, 2-back, 3-back repeated four times, and split into two separate scan runs lasting 8:51 each. Total task time was 17 minutes, 42 seconds. Participants were given a break between scan runs, and asked to indicate when they were ready to proceed.

4.2.2.2.2 Raven’s Progressive Matrices Functional Imaging Task

Raven’s standard set of matrix problems (Raven, 1976) were modified for use in the scanner such that participants could respond using the same scanner-compatible two-button response box utilized for the n-back task. This was achieved within a block-design by presenting participants with each matrix problem for 5000 ms, after which one of the multiple answer options was highlighted with a superimposed red box. Participants were then required to indicate whether this highlighted answer was correct or incorrect via button-press with their index finger or middle finger respectively. See Figure 15, panel B. The matrix problem and highlighted answer option remained on the screen for 8000 ms, and trials were separated by 1000 ms of fixation cross at the centre of the screen. The highlighted answer was the correct answer for exactly half of the trials, and the increased difficulty of ‘good wrong answers’ (i.e. those that closely resemble the actual correct answer) was controlled by ensuring an equal number in each block. The total set of 60 matrix problems that comprise Raven’s Standard Progressive Matrices was split into two sets (for pre- versus post-training comparisons), and further divided into three levels of difficulty: Easy, Medium, and Hard – yielding six sets of 10 problems each. Participants
completed one block of each difficulty level in a randomized order, in a single scanner run. Trials within blocks were also randomized. Task instructions were displayed for 5000 ms before the first block, and each block was preceded by 15 seconds of fixation in order to allow the hemodynamic response to return to baseline. Total task time was eight minutes.

4.2.2.2.3 Lexical Decision Functional Imaging Task

A lexical decision task was included in the protocol as a comparison task that ought not to engage WM or Gf cognitive processes, nor activate the circumscribed fronto-parietal network associated with these tasks. For this block-design task, participants viewed sequences of four letters, and were asked to indicate whether they composed a real word in English via button-press; index finger for ‘yes’, middle finger for ‘no’. See Figure 15, panel C. Four-letter combinations were generated ahead of time using The English Lexicon Project database (Balota et al., 2002), which classifies words and non-words along several empirically derived dimensions, including accuracy of lexical decision (i.e. word versus non-word). For example, ‘blue’ and ‘jume’ have perfect accuracy ratings in the database, indicating that they are relatively easy to distinguish as a word and non-word respectively. In contrast, ‘faze’ and ‘thew’ have considerably lower accuracy ratings reflecting their nature as harder series’ of letters to lexically classify. Participants completed three blocks each of Easy and Hard four-letter combinations, in a single scanner run. Blocks consisted of 30 trials each in which half were real words and half were non-words. Block order was randomized, as were items within each block. Task instructions were displayed for 5000 ms before the first block, and each block was preceded by 15 seconds of fixation in order to allow the hemodynamic response to return to baseline. Total task time was 10 minutes, 45 seconds, making it the longest single task-run of the three included tasks.
4.2.3 MRI Data Acquisition

Whole-brain images were collected for all participants on a 3T General Electric Discovery MR750 system using an 8-channel head coil at the Seaman Family Magnetic Resonance Research Centre at the University of Calgary. Functional echo planar imaging (EPI) data were acquired in an interleaved bottom-up slice order with 40, 3.4 mm thick slices, echo time (TE) = 30 ms, repetition time (TR) = 2500 ms, flip angle = 77°, field of view (FOV) = 22, and matrix = 64 × 64. A total of 424 volumes were collected for the n-back task, 192 volumes for Raven’s Progressive Matrices, and 258 for the lexical decision task. T1-weighted 3D magnetization-prepared rapid gradient-echo (MP-RAGE) anatomical scans were also acquired for each participant to register the functional data (256, 1 mm slices, TE = 3.1 ms, TR = 7.4 ms, inversion time (TI) = 650 ms, FOV= 25.6, matrix = 256 × 256). Additionally, T2-weighted high resolution scans were collected as expanded functional images to aid in registering the functional images to standard space (40, 3.4 mm thick slices, TE = 120 ms, TR = 7500 ms, FOV = 22, and matrix = 256 × 256).

4.2.4 Training Tasks

4.2.4.1 Working Memory Training Program

Three training tasks specifically thought to target WM processes were selected from Lumosity’s broader library of games for the WM training program: 1) ‘Memory Match’ is essentially a visual 2-back task which presents participants with a single line of shapes progressing from right to left across the screen, advancing one position per trial. As the line of randomly ordered shapes passes across the screen, it progresses through two location indicator boxes two positions apart (i.e. with only one space between them). On each trial, participants are asked to indicate via button-press whether the stimuli in the rightmost box matches that in the
leftmost box which contains the stimuli from the rightmost box from two trials previous. These instructions would constitute a simple matching task except that the shapes to the left of the first indicator box become invisible after several correct responses, thereby requiring participants to hold the sequence in memory, and update it after each trial. If participants respond incorrectly, all shapes in the sequence become visible until several subsequent correct responses render these shapes invisible again. 2) ‘Memory Match Overload’ is structured similarly to Memory Match, but positions the indicator boxes two positions apart, thereby making it a more difficult visual 3-back memory task. 3) Finally, ‘Memory Lane’ mimics the structure of the dual n-back task described above. Participants are guided down a virtual street in which each apartment building they pass acts as one trial of the n-back task. For example, at each apartment, a human silhouette appears in one of the windows and auditorily presents a letter. Participants are asked to indicate via button press if either or both the silhouette location and auditorily presented letter are the same as n houses ago. Unlike the previous two training tasks, the Memory Lane game is adaptive. Task difficulty (i.e. n) is increased when participants are successfully completing the task, and decreased when they are not, thereby ‘adapting’ the game to their current skill level. The visual stimuli presentation area (i.e. number of windows per apartment; 2x2 to 3x3), target n are (i.e. number of apartments ago to remember; 1-back to 10-back), and stimuli modality (i.e. visual only vs. both visual and audio) are adjusted according to task performance. See Figure 16, panel A.
Each training session, participants would play each of these games several times in a randomized order. Participants were asked to engage in these tasks for 20 minutes per day, on at least five days per week, for the duration of the six week training period.

4.2.4.2 Processing Speed Training Program

Similar to the WM training program, three tasks thought to target processing speed abilities were selected from Lumosity’s game library to constitute the PS training (i.e. active control) program: 1) ‘Speed Match’ is a speeded visual 1-back task, and requires that participants...
quickly indicate whether a presented shape matches the one that was presented immediately before it. This is a reasonably simple task, however, the emphasis is on improving reaction time over the course of the six week training period. 2) ‘Speed Match Overdrive’ shares a similar structure to Speed Match, but includes a third response option for the currently presented shape to be a ‘partial match’ to that presented directly before it (e.g. matches in colour but not shape, or shape but not colour), thereby making it slightly more challenging. Finally, 3) ‘Spatial Speed Match’ shares the same structure as the Speed Match game, but includes stimuli differing only in spatial orientation. For example, two empty dots and one filled dot might be shown followed by a similar arrangement with the filled dot in a different location, and participants are asked to quickly indicate whether the current pattern matches the previous one or not. See Figure 3, panel B. Importantly, these three processing speed tasks were not directly adaptive in the way that the Memory Lane game was made more or less difficult by altering variables of the game. However, there was an emphasis on constant improvement through reduction of reaction times over the course of training. Consistent with the WM training program, these games were presented in a random order during each training session, and participants in the PS group were asked to engage in these games for approximately 20 minutes per day, on at least five days per week, for the duration of the six week training period.

4.2.5 Data Analysis

4.2.5.1 Behavioural Data Analyses

Potential differences between groups before training on salient demographic and cognitive measures were investigated with independent samples t-tests. Behavioural data collected during scanning sessions was analyzed with a mixed design repeated measures ANOVA modelling examination time (within-subjects; before training versus after training), ×
group assignment (between subjects; WM training versus PS training group) for performance on each of the three scanner tasks (n-back, RSPM, and lexical decision task), as well as for reaction time measures for the latter two. Reaction time data was not analysed for the n-back task, as on exactly half of all trials the correct response was to refrain from responding at all. Questionnaire data collected regarding motivation to complete training, as well as expectations of cognitive improvement as a result of training were similarly analysed with mixed-design repeated measures ANOVAs, though here modelled three time points for the two groups (pre-, during-, and post-training).

4.2.5.2 Neuroimaging Data Analyses

Functional MRI data was analysed with fMRI Expert Analysis Tool (FEAT) and other utilities from FMRIB Software Library (FSL; www.fmrib.ox.ac/fsl) version 5.0.9 for Linux. Preprocessing of functional and anatomical images included reorientation to standard orientation, removal of non-brain tissue, slice-timing correction (Smith, 2002), spatial smoothing (using a 7 mm FWHM Gaussian kernel), grand-mean intensity normalization, and high-pass temporal filtering. Time-series statistical analyses were carried out using FMRIB’s improved linear model (FILM) with local autocorrelation correction (Woolrich 2001). Plots representing head movement revealed less than 2 mm of translation for all participants negating the need for inclusion of motion correction parameters in the preprocessing stage. Functional images were first registered to T2-weighted high resolution images, then to participants’ structural images, and then to standard the MNI avg152 T1-weighted template with non-linear transformations with seven, seven and 12 degrees of freedom respectively using FSL’s FMRIB Linear Image Registration Tool (FLIRT; Jenkinson & Smith, 2001; Jenkinson, Bannister, Brady, & Smith, 2002). Manual inspections were performed after each preprocessing step for each scan run, and
reanalyses were performed where necessary. Group level analyses in FSL were carried out for each group individually, as well as together utilizing a mixed design repeated measures ANOVA which modelled time (within-subjects; before training versus after training), × group assignment (between subjects; WM training versus PS training group) for each pairwise contrast of difficulty levels of the three neuroimaging tasks described above. Figure 17 shows the group level model, and corresponding contrasts of interest.

**Figure 17.** Mixed effects repeated measures ANOVA group level model examining time (within-subjects; pre- vs. post-training), and group (between subjects; working memory training vs. processing speed active control training)

Importantly however, voxels indicating significant time × group interaction effects may represent increased, decreased, or unchanged blood-oxygen-level dependent signal intensity of the WM relative to PS training group. For this reason, all significant time × group interactions were
followed up by investigating the mean intensity changes pre- to post-training for both groups, in only those regions defined by the voxels identified by significant interaction effects. Across all analyses, a Z threshold of 2.3 and cluster threshold of 0.05 were applied (Worsley, 2001), and activation localizations were based on the Harvard-Oxford cortical and subcortical structural atlases provided by FSLView, version 4.0.1. A total of four scanner runs were lost due to problems with data acquisition, ultimately excluding a single task for one participant in the WM group, and all post-training scans for a participant in the PS group. These participants’ behavioural data from the scanner sessions was nonetheless included in analyses of in-scanner task performance and reaction time. Final analyses included 24 participants for the WM training group (n = 23 for the lexical decision task), and 23 participants for the PS active control group.

4.3 Results

4.3.1 Participant Demographics, Cognitive Characteristics, and Personality Variables

The WM training and PS active control groups were equivalent on all variables measured pertaining to cognitive ability and demographic characteristics, including: age \( [t_{47} = -0.38, p = .71] \); balance of males and females \( [\chi^2(1, N = 79) = 0.03, p = .87] \); years of education \( [t_{46} = -0.55, p = .59] \); estimated Full Scale IQ (FSIQ) \( [t_{47} = -0.83, p = .41] \), RAPM performance \( [t_{47} = -1.18, p = .25] \); CCFT performance \( [t_{47} = -1.56, p = .13] \); AOSPAN performance \( [t_{46} = -0.30, p = .77] \); and SDRT performance for both maintenance \( [t_{47} = -1.80, p = .08] \) and maintenance plus manipulation \( [t_{47} = -0.95, p = .35] \) conditions. Groups were also equivalent in terms of scales measuring personality characteristics, including: the Grit scale \( [t_{47} = 0.68, p = .50] \); the Need for Cognition scale \( [t_{47} = 0.13, p = .90] \); current cognitive activities \( [t_{43} = -0.66, p = .51] \), as well as all dimensions of the HEXACO personality inventory. Importantly, the two groups were also equal in terms of their self-rated motivation to complete training \( [t_{47} = -0.39, p = .70] \), and their
expectations of improvement on the training tasks themselves \( t_{47} = 0.35, p = .73 \). Table 5 provides a summary of these group equivalencies before training.

Table 5.

**Participant Characteristics**

<table>
<thead>
<tr>
<th></th>
<th>Working Memory Training Group</th>
<th>Processing Speed Control Group</th>
<th>Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Demographics</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>25</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>30.68 (6.24)</td>
<td>31.33 (5.78)</td>
<td>( t_{47} = -0.38, p = .71 )</td>
</tr>
<tr>
<td>Gender (male/female)</td>
<td>11/14</td>
<td>10/14</td>
<td></td>
</tr>
<tr>
<td>Education (years)</td>
<td>15.24 (2.19)</td>
<td>15.57 (1.93)</td>
<td>( t_{46} = -0.55, p = .59 )</td>
</tr>
<tr>
<td><strong>Cognitive Ability Before</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Training</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WAIS-IV FSIQ</td>
<td>108.24 (15.93)</td>
<td>111.63 (12.34)</td>
<td>( t_{47} = -0.83, p = .41 )</td>
</tr>
<tr>
<td><strong>Personality Factors</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grit score</td>
<td>3.44 (0.69)</td>
<td>3.31 (0.61)</td>
<td>( t_{47} = 0.68, p = .50 )</td>
</tr>
<tr>
<td>Need for Cognition score</td>
<td>69.36 (6.76)</td>
<td>69.06 (9.51)</td>
<td>( t_{47} = 0.13, p = .90 )</td>
</tr>
<tr>
<td>Cognitive Activities (hours/year)</td>
<td>1203.80 (890.50)</td>
<td>1407.52 (1164.60)</td>
<td>( t_{43} = -0.66, p = .51 )</td>
</tr>
<tr>
<td>HEXACO Honesty- Humility</td>
<td>3.74 (0.50)</td>
<td>3.80 (0.54)</td>
<td>( t_{47} = -0.43, p = .69 )</td>
</tr>
<tr>
<td>HEXACO Extraversion</td>
<td>3.00 (0.66)</td>
<td>2.83 (0.80)</td>
<td>( t_{47} = 0.85, p = .40 )</td>
</tr>
<tr>
<td>HEXACO Agreeableness</td>
<td>3.37 (0.53)</td>
<td>3.54 (0.76)</td>
<td>( t_{47} = -0.91, p = .37 )</td>
</tr>
<tr>
<td>HEXACO Conscientiousness</td>
<td>3.84 (0.51)</td>
<td>3.53 (0.64)</td>
<td>( t_{47} = -1.25, p = .22 )</td>
</tr>
<tr>
<td>HEXACO Openness to Experience</td>
<td>3.62 (0.67)</td>
<td>3.93 (0.42)</td>
<td>( t_{47} = -1.91, p = .06 )</td>
</tr>
<tr>
<td><strong>Training Data</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total hours of training</td>
<td>13.49 (4.86)</td>
<td>11.69 (3.03)</td>
<td>( t_{47} = 1.55, p = .13 )</td>
</tr>
<tr>
<td>Pre-training motivation</td>
<td>5.68 (0.87)</td>
<td>5.77 (0.78)</td>
<td>( t_{47} = -0.39, p = .70 )</td>
</tr>
<tr>
<td>Pre-training expectation of improvement</td>
<td>4.52 (0.96)</td>
<td>4.56 (1.10)</td>
<td>( t_{47} = 0.35, p = .73 )</td>
</tr>
</tbody>
</table>

Note. FSIQ, full scale intelligence quotient; WAIS, Wechsler Adult Intelligence Scale
4.3.2 Behavioural Results

4.3.2.1 Training Task Performance and Reaction Time

Participants in both the WM training group and the PS active control group showed improvement on their prescribed training tasks across the training period. Progress in training was measured for each training task by calculating a difference score between performance on their first game, and an average of their last five games. On average, participants in the WM training group achieved an $n$ of 1.80 ($SD = 0.41$) on their first attempt of the Memory Lane game, and progressed to an average $n$ of 4.79 ($SD = 2.08$) across their final five games, yielding a significant mean difference score of 2.99 ($SD = 2.07$), $t(24) = 7.24$, $p < .001$. Additionally, WM training group participants showed increased proficiency on both the Memory Match, and Memory Match Overload games as indicated quantitatively by a greater number of correct matches across their last five games, compared to their first game. Difference scores were significant for both Memory Match $t(24) = 10.15$, $p < .001$, and Memory Match Overload $t(24) = 12.41$, $p < .001$. Participants also attempted a greater number of trials for these two matching games over the course of training, pointing to quicker reaction times, and thereby fitting in a greater number of trials in latter 20-minute training sessions. Training progress in the processing speed active control group was indicated by significant reductions in reaction time across the training period. On average, participants decreased their reaction times by 367.63ms ($SD = 233.11$) on the Spatial Speed Match game, 278.50ms ($SD = 142.27$) on the Speed Match game, and 589.62ms ($SD = 228.14$) when comparing their first game to the mean of their last five games. Difference scores indicated significant reductions in reaction time for each of these difference scores $t(23) = 7.73 - 12.66$, $p < .001$. These training results are displayed graphically in Figure 18.
Figure 18. Mean performance by training game for the working memory training group (A-C), and mean reaction times by training game for the processing speed training group (D-F).
Crucially, the training groups were statistically equivalent in terms of the total amount of time they spent training with their respective online training programs during the training period \( t_{47} = 1.55, p = .13 \).

4.3.2.2 In-Scanner Task Performance Before and After Training

The mixed-design repeated measures ANOVA examining time × group for in-scanner task performance revealed main effects of time for a variety of task conditions, including: 1-back \( F(1,47) = 4.11, p < .05 \), 2-back \( F(1,47) = 44.18, p < .001 \), 3-back \( F(1,47) = 51.89, p < .001 \), and RSPM medium difficulty \( F(1,47) = 9.04, p < .01 \). Main effects of time were also observed for reaction times in the RSPM easy \( F(1,47) = 26.22, p < .001 \) and medium \( F(1,47) = 12.48, p < .001 \) conditions, as well as both easy \( F(1, 47) = 33.71, p < .001 \) and hard \( F(1,47) = 14.61, p < .001 \) difficulty conditions of the lexical decision task. Significant time × group interactions were observed only for performance on the 3-back task \( F(1,47) = 17.04, p < .001 \), the hardest difficulty of the RSPM task \( F(1,47) = 5.88, p = .019 \), and the easy condition of the lexical decision task \( F(1,47) = 5.37, p = .025 \). The time × group interactions for both the 3-back condition and the hard difficulty RSPM task were driven by better task performance by the WM training group compared to the PS active control group at post-training. However, in the case of the RSPM task, the effect was observed not due to an increase in the WM group’s performance after training, but rather a decrease in that of the PS group. The time × group interaction for performance on the easy condition of the Lexical Decision Task was driven by significantly different performance during the first scanning session, before training \( t_{47} = -2.08, p = .043 \).

Thus, the only statistically significant time × group interaction that could reasonably be attributed to the effects of training is that of the most difficult condition (3-back) of the n-back task – with the WM training group outperforming the PS active control group after training. See
Figure 19 for a graphical representation of n-back and RSPM mean performance pre- and post-training.

![Graph A](image1.png)  ![Graph B](image2.png)

**Figure 19.** In-scanner performance on the n-back task (A), and Raven’s Standard Progressive Matrices (B) by group, before and after training. Participants in the WM group performed better on the 3-back condition after training, and participants in the PS group performed worse on the hardest level of the Ravens Progressive Matrices task after training. Error bars represent 95% confidence intervals.

4.3.2.3 Motivation and Expectations for Improvement

Analysis of participants’ expectations for cognitive improvement resulting from WM or PS training, as well as their self-reported motivation to complete the online training itself did not reveal any significant time × group interactions. However, results of the mixed-design repeated
measures ANOVA indicated main effects of time for both motivation \([F(2,84) = 19.40, p < .001]\), and expectations for improvement \([F(2,84) = 5.83, p = .004]\). Participants in both groups indicated a decline in motivation across the training period, but not at significantly different rates. Self-ratings of expectations for improvement followed a U-shaped curve for both groups, with lowest expectations for improvement mid-way through training. See Figure 20.

\[\text{Figure 20. Participants self-reported motivation to complete training (A), and self-reported expectations for cognitive improvement as a result of training (B) on the training tasks throughout the training period. Error bars represent 95\% confidence intervals.}\]

### 4.3.3 Whole Brain fMRI Analyses

#### 4.3.3.1 Dual n-Back Task, by Level of Difficulty

##### 4.3.3.1.1 3-back > 2-back Contrast

Analysis of the greatest WM load condition of the n-back Task versus the intermediate load condition before and after training revealed significant activations for the WM training
group (but not the PS training group), with greater activation following training. Post- relative to pre-training activation was observed in the left lateral occipital cortex, left superior parietal lobule, and left supramarginal gyrus, bilateral precuneus, as well as bilateral areas of the cerebellum. See Table 6, and Figure 21.

Table 6.
Significant whole-brain activations by group, task, and contrast

<table>
<thead>
<tr>
<th>Group, Task, and Contrast</th>
<th>Regions within each cluster</th>
<th>Voxels</th>
<th>Zmax</th>
<th>MNI coordinates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X   Y   Z</td>
</tr>
<tr>
<td><strong>Working Memory Training Group</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dual n-Back Task 3back &gt; 2back</td>
<td>pre &gt; post</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>post &gt; pre</td>
<td>L lateral occipital cortex; R/L precuneus; L superior parietal lobule; L supramarginal gyrus</td>
<td>2433</td>
<td>3.59</td>
</tr>
<tr>
<td></td>
<td></td>
<td>R/L cerebellum</td>
<td>1027</td>
<td>3.42</td>
</tr>
<tr>
<td>3back &gt; 1back</td>
<td>pre &gt; post</td>
<td>R/L frontal poles; R/L paracingulate gyrus; R/L anterior cingulate; L superior frontal gyrus</td>
<td>5903</td>
<td>4.59</td>
</tr>
<tr>
<td></td>
<td></td>
<td>R central opercular cortex; R insular cortex; R orbitofrontal cortex; R temporal pole; R frontal pole; R parahippocampal gyrus; R amygdala</td>
<td>2295</td>
<td>4.02</td>
</tr>
<tr>
<td></td>
<td>post &gt; pre</td>
<td>R/L lateral occipital cortex; R/L superior parietal lobule; R/L precuneus cortex; L postcentral gyrus; L supramarginal gyrus</td>
<td>5494</td>
<td>5.39</td>
</tr>
<tr>
<td></td>
<td></td>
<td>L superior frontal gyrus; L middle frontal gyrus; L inferior frontal gyrus; L paracingulate gyrus; L precentral gyrus; L supplementary motor cortex</td>
<td>2835</td>
<td>4.67</td>
</tr>
<tr>
<td>2back &gt; 1back</td>
<td>pre &gt; post</td>
<td>R/L precuneus; R/L cuneal cortex; R/L posterior cingulate gyrus</td>
<td>2694</td>
<td>4.13</td>
</tr>
<tr>
<td>Task</td>
<td>Group Comparison</td>
<td>Pre &gt; Post</td>
<td>Post &gt; Pre</td>
<td></td>
</tr>
<tr>
<td>-------------------------------------------</td>
<td>------------------</td>
<td>-------------</td>
<td>-------------</td>
<td></td>
</tr>
<tr>
<td>R/L anterior cingulate gyrus; L frontal pole; R/L paracingulate gyrus</td>
<td>2186 4.09 -4 40 20</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R/L superior frontal gyrus; R/L frontal pole</td>
<td>1227 4.47 14 36 54</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>post &gt; pre R superior parietal lobule; R precuneous cortex; R lateral occipital cortex</td>
<td>1088 4.42 16 -56 50</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Raven’s Std. Matrices</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hard &gt; Medium pre &gt; post</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>post &gt; pre L temporal pole; L frontal pole; L inferior frontal gyrus; L orbitofrontal cortex; L parahippocampal gyrus</td>
<td>5079 4.19 -36 36 6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hard &gt; Easy pre &gt; post</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>post &gt; pre R/L lateral occipital cortex; R/L superior parietal lobule; R/L precuneous</td>
<td>6186 4.14 8 -50 68</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Medium &gt; Easy pre &gt; post</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>post &gt; pre L frontal pole; L temporal pole; L orbitofrontal cortex</td>
<td>1361 3.75 -40 46 -20</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lexical Decision Task</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hard &gt; Easy pre &gt; post</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>post &gt; pre R/L frontal pole; R/L frontal medial cortex; R/L anterior cingulate gyrus; R/L paracingulate gyrus</td>
<td>1148 3.48 -24 56 -12</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Processing Speed Training Group</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dual n-Back Task 3back &gt; 2back</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>post &gt; pre</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3back &gt; 1back pre &gt; post</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>post &gt; pre L postcentral gyrus; R/L precuneus; R superior parietal lobule; R lateral occipital cortex; R/L posterior cingulate gyrus; L middle frontal gyrus; L inferior frontal gyrus; L temporal pole; R temporal occipital fusiform gyrus; R caudate; R/L</td>
<td>12086 4.49 46 -54 -18</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Task</td>
<td>Pre vs. Post</td>
<td>Brain Regions</td>
<td>T-Value</td>
<td>M1</td>
</tr>
<tr>
<td>----------------------------------------------------------------------</td>
<td>--------------</td>
<td>----------------------------------------------------</td>
<td>---------</td>
<td>----</td>
</tr>
<tr>
<td>2-back &gt; 1-back</td>
<td>pre &gt; post</td>
<td>-</td>
<td></td>
<td>1631</td>
</tr>
<tr>
<td>post &gt; pre</td>
<td>R superior frontal gyrus; R middle frontal gyrus; R/L paracingulate gyrus; R anterior cingulate gyrus</td>
<td>2336</td>
<td>3.96</td>
<td>46</td>
</tr>
<tr>
<td>Raven’s Std. Matrices</td>
<td>pre &gt; post</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hard &gt; Med</td>
<td>post &gt; pre</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hard &gt; Easy</td>
<td>pre &gt; post</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Medium &gt; Easy</td>
<td>post &gt; pre</td>
<td>R lateral occipital cortex; R superior parietal lobule; R supramarginal gyrus</td>
<td>4376</td>
<td>5.03</td>
</tr>
<tr>
<td>Lexical Decision Task</td>
<td>pre &gt; post</td>
<td>R angular gyrus; R lateral occipital cortex</td>
<td>1635</td>
<td>3.84</td>
</tr>
<tr>
<td></td>
<td>post &gt; pre</td>
<td>L lateral occipital cortex; L angular gyrus</td>
<td>1003</td>
<td>3.66</td>
</tr>
<tr>
<td>Hard &gt; Easy</td>
<td>pre &gt; post</td>
<td>R/L lateral occipital cortex; R/L precuneus; R/L superior parietal lobule</td>
<td>1148</td>
<td>3.48</td>
</tr>
</tbody>
</table>
For this contrast, clusters with significant time × group interaction terms were identified in the bilateral paracingulate gyrus, bilateral anterior cingulate gyrus, and bilateral frontal poles. See Table 7, and Figure 22 panel A. Visual inspection of extracted signal intensity values indicate
that this interaction was driven by reduced relative activation for the 3-back versus 2-back task conditions in these areas for the WM group after training. See Figure 22 panel B.

Table 7.

**Significant whole-brain activations identified by significant time × group interaction terms**

<table>
<thead>
<tr>
<th>Task and Contrast</th>
<th>Regions within each cluster</th>
<th>Voxels</th>
<th>Zmax</th>
<th>MNI coordinates</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Dual n-Back Task</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3back &gt; 2back WM diff &gt; PS diff</td>
<td>R/L paracingulate gyrus; R/L anterior cingulate gyrus; R/L frontal pole</td>
<td>2230</td>
<td>3.53</td>
<td>10  52  6</td>
</tr>
<tr>
<td>3back &gt; 1back WM diff &gt; PS diff</td>
<td>R/L superior frontal gyrus; R/L paracingulate gyrus; R/L cingulate gyrus</td>
<td>5578</td>
<td>4.38</td>
<td>18  30  56</td>
</tr>
<tr>
<td></td>
<td>R insular cortex; R temporal pole; R parahippocampal gyrus</td>
<td>3199</td>
<td>4.69</td>
<td>42  -6  -8</td>
</tr>
<tr>
<td></td>
<td>L insular cortex; L temporal pole; L parahippocampal gyrus</td>
<td>2220</td>
<td>3.97</td>
<td>-38  2  -24</td>
</tr>
<tr>
<td></td>
<td>R/L posterior cingulate gyrus;</td>
<td>1680</td>
<td>3.89</td>
<td>14  -44  32</td>
</tr>
<tr>
<td>2back &gt; 1back WM diff &gt; PS diff</td>
<td>R insular cortex; R middle temporal gyrus; R angular gyrus; R supramarginal gyrus</td>
<td>2595</td>
<td>4.46</td>
<td>36  -2  -6</td>
</tr>
<tr>
<td></td>
<td>R/L posterior cingulate gyrus</td>
<td>1848</td>
<td>4.03</td>
<td>-2  -30  26</td>
</tr>
<tr>
<td></td>
<td>L postcentral gyrus; L insular cortex; L temporal pole</td>
<td>1082</td>
<td>3.40</td>
<td>-44  22  -22</td>
</tr>
<tr>
<td><strong>Raven’s Std. Matrices</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hard &gt; Medium PS diff &gt; WM diff</td>
<td>L inferior frontal gyrus; L frontal pole</td>
<td>1325</td>
<td>3.90</td>
<td>-36  36  6</td>
</tr>
<tr>
<td>L precentral gyrus; L postcentral gyrus; L superior frontal gyrus</td>
<td>1180</td>
<td>3.88</td>
<td>-46  -22  54</td>
<td></td>
</tr>
</tbody>
</table>
Figure 22. (A) Significant whole-brain time × group interaction activations for the 3-back > 2-back contrast of the dual n-back task. (B) Mean signal intensity in contrast of parameter estimate (COPE) values by time and group for the voxels identified in panel A. Error bars represent 95% confidence intervals.
4.3.3.1.2 3-back > 1-back Contrast

Analysis of the greatest WM load condition of the n-back versus the lowest WM load condition before and after training revealed significant activations for the WM training group for both pre > post and post > pre training contrasts. Pre- relative to post-training activations included: bilateral frontal poles, bilateral anterior cingulate gyrus, bilateral paracingulate gyrus, left superior frontal gyrus, right central opercular cortex, right insular cortex, right orbitofrontal cortex, right temporal pole, right frontal pole, right parahippocampal gyrus, and right amygdala. Post- relative to pre-training activation was observed for the WM training group in bilateral lateral occipital cortex, bilateral superior parietal lobule, bilateral precuneus cortex, left postcentral gyrus, left supramarginal gyrus, left superior frontal gyrus, left middle frontal gyrus, left inferior frontal gyrus, left paracingulate gyrus, left precentral gyrus, and left supplementary motor cortex. The PS training group only showed activation for post- relative to pre-training, and was observed in a wide variety of dispersed regions across the frontal, parietal, temporal and occipital lobes, and also the cerebellum. See Table 6, and Figure 23.
Figure 23. Significant whole brain activations for the 3-back > 1-back contrast of the dual n-back task.

Clusters indicating significant time × group interaction effects were revealed in a wide variety of areas including bilateral superior frontal gyrus, bilateral paracingulate gyrus, bilateral cingulate gyrus, bilateral insular cortex, bilateral temporal poles, bilateral parahippocampal gyrus, as well as bilateral posterior cingulate cortex. See Table 6, and Figure 24 panel A. Visual inspection of
Extracted signal intensity values again indicated reduced relative activation for the 3-back versus 1-back task conditions in these areas for the WM group after training. See Figure 24 panel B.

**Figure 24.** (A) Significant whole-brain time × group interaction activations for the 3-back > 1-back contrast of the dual n-back task. (B) Mean signal intensity in contrast of parameter estimate (COPE) values by time and group for the voxels identified in panel A. Error bars represent 95% confidence intervals.
4.3.3.1.3  2-back > 1-back Contrast

Analysis of the intermediate WM load condition of the n-back versus the lowest WM load condition before and after training revealed significant activations for the WM training group for both pre > post and post > pre training contrasts. Pre- relative to post-training activations for the WM group included bilateral precuneus, bilateral cuneus, bilateral anterior cingulate gyrus, bilateral frontal poles, and bilateral paracingulate gyrus. Post- relative to pre-training activations were observed for the WM training group on this contrast in the right superior parietal lobule, right precuneus, and right lateral occipital cortex. By comparison, the PS active control group showed only post- relative to pre-training activation in a wide variety of areas including: left postcentral gyrus, bilateral precuneus, right superior parietal lobule, right lateral occipital cortex, bilateral posterior cingulate gyrus, left middle frontal gyrus, left inferior frontal gyrus, left temporal pole, right temporal occipital fusiform gyrus, right caudate, bilateral pallidum, bilateral thalamus, bilateral cerebellum, bilateral superior frontal gyrus, right middle frontal gyrus, bilateral paracingulate gyrus, and right anterior cingulate gyrus. See Table 6, and Figure 25.
Figure 25. Significant whole brain activations for the 2-back > 1-back contrast of the dual n-back task.

For this contrast, clusters with significant time × group interaction terms were identified in the right insular cortex, right middle temporal gyrus, right angular gyrus, right supramarginal gyrus, bilateral posterior cingulate gyrus, as well as left postcentral gyrus, left insular cortex, and left temporal pole. See Table 6, and Figure 26 panel A. Visual inspection of extracted signal intensity
values again indicated reduced relative activation for the 2-back versus 1-back task conditions in these areas for the WM group after training. See Figure 26 panel B.

Figure 26. (A) Significant whole-brain time × group interaction activations for the 2-back > 1-back contrast of the dual n-back task. (B) Mean signal intensity in contrast of parameter estimate (COPE) values by time and group for the voxels identified in panel A. Error bars represent 95% confidence intervals.
4.3.3.2 Raven’s Standard Progressive Matrices, by Level of Difficulty

4.3.3.2.1 Hard > Medium Contrast

Comparing the two most difficult levels of Raven’s progressive matrices yielded significant training related activations only for the WM training group, and at post- relative to pre-training. Activation included the left temporal pole, frontal pole, left inferior frontal gyrus, left orbitofrontal cortex, and left parahippocampal gyrus. See Table 6, and Figure 27.
Figure 27. Significant whole brain activations for the Hard > Medium contrast of the Raven’s Standard Progressive Matrices task.

Clusters indicating significant time × group interaction effects were revealed in left inferior frontal gyrus, left frontal pole, left precentral gyrus, left postcentral gyrus, and left superior frontal gyrus. See Table 6, and Figure 28 panel A. Extracted signal intensities indicated that the time × group interaction for this contrast was driven by increased relative activation for the Hard
versus Medium task conditions in the above areas for the WM group after training. See Figure 28, panel B.

Figure 28. (A) Significant whole-brain time × group interaction activations for the Hard > Medium contrast of the Raven’s Progressive Matrices task. (B) Mean signal intensity in contrast of parameter estimate (COPE) values by time and group for the voxels identified in panel A. Error bars represent 95% confidence intervals.
4.3.3.2.2 Hard > Easy Contrast

Comparing the most and least difficult conditions of the RSPM task indicated significant activations for both groups only at post- relative to pre-training. Activations for the WM group included bilateral lateral occipital cortex, bilateral superior parietal lobule, bilateral precuneus. The PS group showed activation for the post > pre contrast in exclusively right-lateralized areas including the right superior frontal gyrus, right middle frontal gyrus, and right precentral gyrus. See Table 6, and Figure 29. No clusters evinced time × group interaction effects for this contrast.
4.3.3.2.3 Medium > Easy Contrast

Analysis of the easiest two conditions of Raven’s progressive matrices task revealed significant activations for both groups, but only in the pre- relative to post-training contrast. The WM group showed activation in the left frontal pole, left temporal pole, and left orbitofrontal...
cortex. The PS training group showed activation for the pre > post contrast in the bilateral angular gyrus, and the bilateral lateral occipital cortex. See Table 6, and Figure 30. No clusters evinced time × group interaction effects for this contrast.

<table>
<thead>
<tr>
<th></th>
<th>pre &gt; post</th>
<th>post &gt; pre</th>
</tr>
</thead>
<tbody>
<tr>
<td>WM</td>
<td></td>
<td>----</td>
</tr>
<tr>
<td>Group</td>
<td></td>
<td>----</td>
</tr>
<tr>
<td>PS</td>
<td></td>
<td>----</td>
</tr>
</tbody>
</table>

*Figure 30*. Significant whole brain activations for the Medium > Easy contrast of the Raven’s Standard Progressive Matrices task.
4.3.3.3 Lexical Decision Task, by Level of Difficulty

4.3.3.3.1 Hard > Easy Contrast

Comparing the two difficulty levels of the lexical decision task revealed significant activations for both the WM and PS groups, but only for the pre- relative to post-training contrast. The WM group showed activation in exclusively frontal regions including the bilateral frontal poles, bilateral frontal medial cortex, bilateral anterior cingulate gyrus, and bilateral paracingulate cortex. In contrast, the PS group showed activation in exclusively posterior regions including bilateral superior parietal lobule, bilateral lateral occipital cortex, and bilateral precuneus. See Table 6, and Figure 31. No clusters evinced time × group interaction effects for this contrast.
Figure 31. Significant whole brain activations for the Hard > Easy contrast of the lexical decision task.

4.4 Discussion

The purpose of the present study was to investigate the neural correlates of WM training, as well as its purported transfer to abilities in other cognitive domains in healthy young adults. While the behavioural study of WM training has received much empirical attention over the last
several years, very few studies have approached the central question of this literature from the perspective of neuroimaging. That is, if WM training does confer higher ability on untrained tasks in domains outside of WM, we should be able to discern clear and unambiguous time × group interactions in the networks known to support these kinds of cognition. We investigated this proposition in a single-blind randomized controlled six week trial of online WM training compared to an active control group, in healthy adults aged 18-40 recruited from the community. To our knowledge, this is the first trial to utilize scanner tasks representing both the domain of cognitive training (i.e. WM), as well as the domain to which that training is purported to transfer to (i.e. Gf), utilizing both an active training group as well as an active control condition.

Present results are generally consistent with the patterns observed in the few existing neuroimaging investigations of WM training trials: 1) participants improved their performance on the training tasks to which they were assigned; 2) participants improved their performance on several tasks that were not practiced over the training period (though importantly, not at significantly different rates between WM training and PS active control groups), and; 3) patterns of activation on WM tasks (but not Gf tasks) were observed to decrease in key WM network areas in response to WM training (but not PS training).

4.4.1 Behavioural Results

First, participants in both the WM training group and PS active control conditions were observed to significantly improve their performance on all three of the tasks which they were assigned to train with for approximately 12 hours over the course of the six week training period. Members of the WM training group demonstrated the ability to commit increased numbers of stimuli to memory in the service of n-back matching tasks, and were also observed to perform these tasks more quickly over the training period. Similarly, members of the PS group reduced
their reaction times on several similar matching tasks which did not require committing stimuli to memory. Interestingly, average reaction times for all three games constituting the PS training program continued a downward trend, even towards the end of training, indicating both continued active engagement with the tasks as well as the potential for even greater performance gains with continued training. This is an important quality for the active control condition, as some researchers have suggested that the adaptive nature of the dual n-back task often used for training is not well controlled for by non-adaptive tasks (see Brehmer, Westerberg, & Bäckman, 2012; von Bastian & Eschen, 2016). Contrary to this notion, PS group performance actually continued to improve over the course of training, whereas our WM group appeared to plateau or even decline over the latter portion of the training period. This pattern of WM training (specifically with n-back) meeting with limiting returns, or reversing in latter sessions has been observed previously in the literature (see Sprenger et al., 2013 – Figure 2, pp. 655).

Despite this plateau in performance on training tasks themselves, the WM group showed improved task performance over the PS group on the most difficult condition of the dual n-back task while in the MRI scanner after training. This significant time × group interaction demonstrates the near-transfer effect of WM training to similar tasks that is often observed in the literature. Analysis of the in-scanner behavioural data for our transfer task (RSPM) also indicated a significant time × group interaction, though due to unexpectedly low scores in the PS group versus their pre-training performances. The reasons for this substantial drop in task performance in the PS group are not clear. Reaction times for this task did not vary between groups, suggesting that poorer performance was not driven by some training-specific mechanism such as an increased pressure to respond quickly in the PS group. Regardless, visual inspection of the data has been emphasized in these cases in order to ensure that significant interaction effects can
sensibly be attributed to the effects of training, or (as in this case) not (Redick, 2015, Wagenmakers, 2015).

4.4.2 Neuroimaging Results

Load-dependant activation patterns at pre-training MRI scanning were largely consistent with those observed in the literature for WM tasks (Owen et al., 2005; Rottschy et al., 2012), as well as Gf tasks (Prabhakaran, Smith, Desmond, Glover, & Gabrieli; Christoff et al. 2001; Kroger, Sabb, Bookheimer, Cohen, & Holyoak, 2002; Perfetti, Saggino, Ferretti, Caulo, Romani, & Onofrj, 2009). The mean effects of each of the pairwise harder > easier contrasts for each of our three tasks collapsed across both groups is described in Clark et al. (in preparation¹).

In general, increasing task WM task demands revealed the expected pattern of activation increases across a wide variety of fronto-parietal areas including the middle and superior frontal gyri, frontal poles, frontal orbital cortex, insular cortex, bilateral paracingulate gyrus, as well as bilateral precuneus, superior parietal lobule, supramarginal gyrus, and angular gyrus. Left caudate and bilateral areas of the cerebellum were also implicated. Similarly, mean load-dependent activation increases for the RSPM task of Gf were observed in right inferior and superior frontal gyri, as well as left lateral occipital cortex, left superior parietal lobule, and left occipital pole. Interestingly however, roughly similar areas of mean fronto-parietal activation were also observed in our lexical decision scanner task which represents a wholly different realm of cognition (i.e. vocabulary, crystallized intelligence), and is not thought to significantly engage WM processes. These results are consistent with the notion that any new sufficiently difficult task may engage a core neural network involved in WM and attention more generally (Fedorenko, Duncan, & Kanwisher, 2013). For this reason, this network has garnered a large number of eponyms in the research literature describing its centrality in virtually all non-routine
task-based cognitive endeavours. These titles include: the attention and working memory system (Cabeza & Nyberg, 2000), the cognitive control network (Cole & Schneider, 2007), the task-positive network (Fox, Snyder, Vincent, Corbetta, Essen, & Raichle, 2005), the executive control network, or dorsal attention network (Yeo et al. 2011), the core executive (discussed in Rottschy et al. 2012), and the multiple demand network (Duncan, 2010).

So while our tasks reliably and predictably engaged this core multiple demand network, the key question of the present investigation is how activation of that network is impacted by intensive and adaptive training of WM compared to a similar control condition. Indeed, if this network is utilized by a wide range of tasks encompassing WM and attention, the thinking is that functional changes in this network could be sufficient to explain the elusive effects of WM training to seemingly far-flung domains of cognition like Gf. On this question, results indicated decreases in fronto-parietal activation in the WM group relative to the PS control group for all harder > easier contrasts of the dual n-back task. Specifically implicated areas of activation decrease included bilateral paracingulate, anterior cingulate, and frontal poles, as well as bilateral superior frontal gyrus, insular cortex, temporal poles, parahippocampal gyri, as well as right supramarginal and angular gyri. This is consistent with previous investigations (Hemipel et al., 2004; Schneiders et al., 2011; Schneiders et al., 2012; Thompson et al., 2016), and confirms our first hypothesis. Results of analyses examining the WM group alone suggest reductions in activation in frontal regions, and increases in parietal regions across the training period in response to the n-back task (see Figures 21, 23, 25). This is consistent with what would be predicted by the P-FIT model of intelligence (Jung & Haier, 2007), as well as the small literature suggesting that individuals of higher intelligence tend to rely to a greater extent on parietal areas while completing complex tasks rather than frontal areas when compared to those of lower
ability (Jaušovec & Jaušovec, 2012). Present results may indicate a reduced requirement for attentional control for the WM trained group while completing the n-back task after training, or as Thompson and colleagues (2016) note: “reduce[d] demand on higher levels of a hierarchical system that supports the maintenance and updating of active WM traces” (pp. 11). This training-induced reduction in activation might be interpreted as thereby mimicking the neural approach of a naturally higher ability individual.

These reductions in activation following training also suggest that the training period was sufficiently lengthy to facilitate this adaptation. Following the results of earlier neuroimaging studies, we might have expected activation increases in the WM core network with a shorter duration of training (e.g. Hempel et al., 2004), or also in cases of similar length but with more difficult post-training cognitive tests (e.g. 4- or 5-back conditions; Buschkuehl et al., 2014). Additionally, we might have expected further reductions in activation in the WM training group relative to the PS group if the training task was even more similar to that of the in-scanner dual n-back task, rather similar to Schneider and colleagues (2011, 2012) notion of intra- versus cross-modal training. Thus, consistent with the few existing studies which have examined neural activation changes in response to cognitive training, the present trial adds to the weight of evidence suggesting remarkably plastic capabilities of the human brain in response to repetitive practice with a complex task.

However, the main question of interest here was not the neural mechanisms of practice effects for a WM task (or a very similar task), but rather the neural activations associated with the purported behavioural far-transfer of that increased ability to tasks in the domain of Gf. To this question, our results implicate an exclusively left sided grouping of voxels in frontal lobe and postcentral gyrus as representing a time × group interaction effect when comparing the two
most difficult levels of the RSPM task. However, contrary to the activation reductions observed in the WM group for the dual n-back task, here further inspection of the mean signal intensity in voxels implicated in the interaction effect indicates relative activation increases in the WM group compared to the PS group, as well as unexpected and significant differences between groups at pre-training. While it might be reasonable to suggest that this increase in neural activation represents a kind of training-induced acquired aptitude for the most difficult level of the task in the WM trained group, consideration of the behavioural performance of the groups would suggest otherwise. Unfortunately, the unexpected task performance decreases on the part of the PS group at post-training scanning, to a large extent undermine the conclusions that can be drawn from statistical comparisons of activation patterns between the groups. That is, it is not possible to discern on the basis of the observed data whether the interaction effect was driven by reductions in the PS group as a result of lack of cognitive effort leading to poorer performance, or activation increases in the WM group due to some process precipitated by training. Whichever the case may be, our second hypothesis (post-training reductions in fronto-parietal activation for difficult levels of our transfer task) is not supported by the present data. Conversely however, and consistent with our third hypothesis, training was not associated with any reliable activation changes between groups for the lexical decision task which represents a cognitive domain (i.e. vocabulary, or crystallized intelligence) that theoretically ought not to have been affected by training at all.

4.4.3 Implications for the WM Training Literature

So what do these findings tell us about the nature of WM, neuroplasticity, intelligence, and the WM training literature more broadly? Several researchers have suggested that these kinds of training-induced reductions in activation represent a kind of acquired neural efficiency
for the task after training (Buschkuehl et al., 2014), however, this is difficult to comment on conclusively because the neural mechanisms of WM itself are not yet fully understood, nor are the neuroplastic processes by which they might change with repeated practice or training. For these reasons, Constantinidis and Klingberg (2016) stress that while fMRI studies can point to areas of the brain that are implicated in WM training-related changes, they offer little in terms providing insight about underlying cellular mechanisms. For example, activation decreases may represent lower rates of firing by cells in the network, or perhaps a decrease in the number of cells responding at the same firing rate to WM task demands. Proposed mechanisms of neuroplasticity include the activation and suppression of individual genes, changes in the strength of synapses, changes in the number of dendritic spines, changes in the arborisation of axons, and changes in axon myelination. Constantinidis and Klingberg (2016) wisely note the continued challenges in relying on neuroscience to inform psychology about previously unknown psychological mechanisms of learning, such as WM training.

However, the present results do provide some clarity on the WM training literature, particularly in terms of the purported neural mechanisms of far-transfer of WM training. Before very recently, the neural effects of cognitive training were not known, and were generally discussed only in terms of taxonomies of possibilities (e.g. activation increases, or decreases, or redistribution, or reorganisation; Kelly and Garavan, 2005, Kelly, Foxe, & Garavan 2006). The present results are consistent with the small number of recent studies which have generally supported the second of these possibilities in response to WM training: activation decreases (Schneiders et al., 2011; Schneiders et al., 2012; Thompson et al., 2016). Additional evidence supports the existence of a complex time course with repeated task practice, with these activation decreases following earlier increases in response to training (Hempel et al, 2004; Olesen et al.,
2004; Buschkuehl et al., 2014). Taken together these studies support the link between better task performance, and less activation of the fronto-parietal core WM network during performance of a trained task. That is, repeated practice with a task has the effect of making its requirements more automatic over the course of training, allowing a relaxation of previously required executive oversight. Rather like learning any new complex skill, practicing with the dual n-back task reduces the requirement for effortful attention to its various sub-components (e.g. switching between monitoring, maintaining, and encoding successive stimuli in two modalities, while also inhibiting response to lure trials in both these modalities). Given this, it seems reasonable that if facilitated performance were to occur for a transferred-to task of Gf, we would also expect reductions in activation for these tasks. This is corroborated by the cognitive theory underpinning far-transfer of WM training which suggests that Gf is facilitated by alleviating the shared capacity constraint between tasks like dual n-back and RSPM which is the shared demand for attention (Halford, Cowan, & Andrews, 2007). Now that we know how that alleviation of the demand for attention reveals itself in terms of neural activation patterns for trained WM tasks, we can reasonably assume that it should be the same for any proposed transfer tasks as well. On the basis of this model we would not, for example, expect WM training to be associated with reduced fronto-parietal activations for WM tasks, but increased activation in these same areas for tasks of Gf. Indeed, from a neural perspective, it could be argued that Gf tasks like RSPM present such similar sub-task component problems as the dual n-back task, that it is simply a variety of visuospatial WM task (Clark et al., in preparation¹).

Unfortunately, the present results fall short of being able to comment on time × group interaction effects in the RSPM transfer task due to the unexpected drop in task performance of the PS group at post-training testing discussed above. We would also not have expected
neuroimaging indications of far-transfer given that cognitive outcomes of training revealed no convincing evidence of transfer to either WM capacity, or more distal domains of cognition such as $Gf$ (Clark et al., in preparation). However, by providing further support for the now familiar pattern of fronto-parietal activation decreases in response to WM training, we can help formulate an important prediction for future trials investigating the neural mechanisms of WM training and transfer: The only sensible pattern of neuroimaging results that would help support the notion of behavioural far-transfer of WM training to tasks of $Gf$ is fronto-parietal decreases in both the training and transfer tasks.

4.4.4 Limitations

Limitations of the current trial include high participant attrition, and in-scanner training and transfer tasks that do not match their out-of-scanner equivalents in several important ways. First, regarding participant attrition, while only 7, and 8 participants withdrew from the Lumosity WM training and PS active control groups (or abandoned their assigned training plan) after randomization respectively, this drop-out represents a relatively large proportion of the total groups (7/32 = 21.89% for the WM group, and 8/32 = 25% for the PS group). High attrition in our sample may speak to any number of factors about the tolerability of the Lumosity interventions, and unfortunately leaves the current results open to speculation about potential systematic differences between participants who completed the training and those that did not. Second, both the dual n-back task and the RSPM tasks were modified for use in the scanner in ways that may have potentially impacted participants’ cognitive approach to them. Specifically, while the dual n-back task is generally implemented in an adaptive fashion, here we advanced participants through 1-back, 2-back, and 3-back condition blocks in order regardless of task performance. We did this to avoid any potential confusion over which condition participants
were completing, though doing so may have precipitated some ordering or expectancy effects. Additionally, the RSPM task was implemented in a temporally imprecise block design which allowed an 8000 ms window of time for participants to choose the correct answer. Average reaction times to this task were between 2000 and 3000 ms, leading to a large proportion of block time representing non-task related activity. Our in-scanner RSPM task additionally failed to discriminate between difficulty conditions on some metrics (e.g. performance accuracy, reaction time; discussed in Clark et al., in preparation) suggesting that splitting the original RSPM question set into two rather than three levels of difficulty may have yielded more distinct comparisons.

4.4.5 Conclusions

The above limitations aside, the current trial adds meaningfully to the body of work on WM training in several important ways. First, it replicates and affirms patterns of neural activation decrease for WM tasks in the fronto-parietal network in response to WM training. Second, to our knowledge it is the first study to investigate neural activation patterns in both training and transfer tasks, and only the second study to examine these patterns in an active control condition as well as a WM training group. Contrary to predictions, WM training did not reliably precipitate reductions in fronto-parietal activation for transfer tasks of Gf. While this null result is consistent with the null cognitive outcomes from our WM training trial (Clark et al., in preparation) it also solidifies the prediction that far-transfer of WM training to Gf ought to be associated with similar reductions in fronto-parietal activation for the transfer tasks that are observed for the purported training task. Future research might use this neuroimaging-based prediction to investigate neuroplastic effects of training and transfer tasks in alternative populations for which WM is impaired, and/or WM training has shown more promise in
behavioural trials, including in various mental disorders (Ansari, 2015), acquired brain injury (Weicker et al., 2016), or healthy older adults (Karbach & Verhaeghen, 2014; Corbett et al., 2015).
4.5 References


186


doi:10.1016/j.bandc.2012.02.007


doi:10.1093/cercor/12.5.477


doi:10.1016/j.intell.2008.08.004


Chapter Five: General Discussion

The present series of studies sought straightforward answers to basic unresolved questions in the working memory (WM) training literature - namely, 1) whether or not working memory (WM) training ‘works’; as well as 2) how the transfer of training might occur on a neural level, if it were indeed found to transfer at all.

5.1 Does Working Memory Training Work?

On this first question, the cognitive outcomes of our investigation strongly support an answer in the negative, at least for healthy young adults recruited from the community, who trained online at home for roughly six weeks with Lumosity’s WM-specific games. Our results unambiguously support the null hypothesis not only for the far-transfer of WM training to two separate measures of fluid intelligence (Gf), but also for the near-transfer of WM training to a number of measures of WM capacity. In the theory of Halford, Cowan, and Andrews (2007), near-transfer to WM capacity is thought of as a kind of pre-requisite for far-transfer. In this sense, it is perhaps not surprising that we found no evidence of far-transfer. Indeed finding far-transfer in the absence of near-transfer would not only have been an unexpected result, it would have raised problems in reconciling our data with Halford et al.’s (2007) proposed mechanism of training transfer itself. Given the extremely split nature of the literature with respect to WM training efficacy, our results are consistent with a large number of studies, and inconsistent with a large number of other studies. While the addition of our single experimental trial to this overall literature cannot hope to resolve the question of efficacy, comparison of study methodological characteristics is telling. As discussed in our second manuscript (Clark, Lawlor-Savage, & Goghari, in preparation\textsuperscript{2}), null results for far-transfer effects are commonplace in studies with both 1) active control groups, and; 2) greater than 20 participants per group. When considering
the wider literature with less ideal methodological characteristics, the effects are still less than desirable, or, as Makin (2016) notes: “…the effect of working memory training on fluid intelligence lies somewhere between zero and very small, depending on whose analysis you trust” (pp. S11)

Given this across-the-board null result for transfer of training, the question is of course why? Why didn’t we observe improved WM capacity or Gf as many other studies have? There are three main possibilities here. First, it could be the case that our training tasks were ineffective at increasing WM capacity, which then all but prevented further far-transfer effects. This kind of near-transfer effect is common in the literature, though is generally thought of as an unexciting finding because it is so close to the domain being trained. The ineffectiveness of our online Lumosity training tasks remains a possibility, however, several indications suggest otherwise. For example, participants in the final analysis remained engaged in the training tasks throughout training, and were shown to statistically improve their scores on training when comparing their first task effort to the average of their final five. As discussed in our second manuscript, this failure of near-transfer, while unexpected, is not a novel finding in the literature (see Redick et al., 2013). The second possible explanation for our failure to find near- or far-transfer effects of training is our rather small sample sizes. While we did include just over the suggested minimum number of participants per group (Simmons, Nelson, & Simonsohn, 2011), power analyses revealed a much higher required number for the desired power of 0.8. Thus, if as Makin (2016) and recent meta-analyses (Melby-Lervåg, Redick, & Hulme, 2016) suggest, the behavioural effects of WM training are miniscule, the fact is that our trial had a poor chance of detecting it in our small sample. The final possible explanation for the failure of our trial to find training effects is that there simply is no effect to find, at least in our population. For example, it may be the case
that WM training helps to facilitate WM and its cognitive correlates in a remedial capacity only (e.g. in mental illness, acquired brain injury, or age-related cognitive decline), and not in a facilitative one for healthy young adults at or near their peak of cognitive ability. On this point, our sample was on average highly educated, with an estimated average of close to 16 years of education. Of course, it may also be the case that WM training does not remediate WM or Gf in these capacities either, though current WM training research in clinical populations appears to hold slightly more promise than that in healthy adults (Ansari, 2015; Weicker, Villringer, & Thöne-Otto, 2016; Corbett et al., 2015).

Importantly, while the above variables may individually or jointly account for the observed null cognitive effects, several other important potential moderators can be ruled out as contributors. These include a range of intra-personal characteristics that have variously been thought to moderate treatment effect, including: motivation to complete training, expectations of improvement from training, major personality traits (Ashton & Lee, 2009), ‘grit’ (i.e. long-term determination; Duckworth & Quinn, 2009), need for cognition (Cacioppo, Petty, & Kao, 1984), as well as current cognitive activities (Eskes, et al., 2010). It remains possible however, that the entire group of participants studied here demonstrates a restricted range on one or more of these variables, making them a non-representative sample of the greater population that may stand to benefit from WM training. For example, our rather highly educated sample may have higher than average levels of conscientiousness, or grit, or may engage in a greater number of cognitively challenging activities that the general population. If this were the case, these characteristics on their own might account for the observed null effects of WM training. Each of these is an empirical question for future research. Other open empirical questions include whether or not substantial transfer of training was exhibited by any of the participants in our sample (regardless
of group averages), and whether combined training approaches or multifaceted interventions may be more effective in facilitating cognitive outcomes of interest in healthy young adults (see Mrazek, Mooneyham, Mrazek, & Schooler, 2016 for a particularly audacious trial in this regard). Additionally, results of large multi-site training studies with standardized training delivery may help resolve the current problems associated with inter-study methodological variability and small sample sizes (see von Bastian, Locher & Ruflin, 2012 for an early example).

Finally, it warrants mention here that ‘placebo effects’ resulting from increased motivation for training, or expectation of improvement from training can also be ruled out as contributors to the null effect in our trial. Both of these constructs were measured, and found to be equal across both the groups that engaged in online training. While this is a particular strength of our trial, it is interesting to note that manipulation of these factors themselves, (i.e. actively seeking to elicit placebo effects) has in at least one case produced a facilitative effect on outcome measures of $G_f$ that WM training itself has often failed to (Foroughi, Monfort, Paczynski, McKnight, & Greenwood, 2016). The authors discuss their findings in terms of increasing awareness of potentially suggestive recruitment methods for WM training trials, however in other domains of clinical research this suggestibility is itself the intervention of interest. As an example, consider the work of Levy (1996) who showed that memory performance in older adults could be improved by merely implicitly influencing their stereotypes about cognition in aging. In this sense, recent scientific investigations of WM training may have been (or continue to be) influenced by the implicit expectation effects driven by the marketing of the industry itself. In other words, the brain training industry may have been running on the steam of its own bluster for the past several years, with their own advertising sowing the seeds of placebo to be harvested later as evidence in methodologically poor empirical trials. While this certainly is a
potential confounding variable to be separated from any ‘true effects’ of WM training, the powerful effects of placebo ought to also be kept in mind as potentially meaningful targets of intervention as well.

5.2 How Reasonable are the Proposed Mechanisms for Transfer of WM Training?

Despite the null cognitive results of the present trial, our neuroimaging investigations revealed several interesting findings that help to clarify a number of claims in the WM training literature, and also help to formulate predictions for future neuroimaging research. Specifically, by exploring load dependent patterns of neural activation in response to training (i.e. dual n-back) and transfer tasks (i.e. Raven’s Matrices) typically utilized in the WM training literature, we are uniquely situated to comment on the feasibility of proposed neural mechanisms of far-transfer. Indeed, to our knowledge, our first manuscript (Clark, Lawlor-Savage, & Goghari, in preparation) is the first to examine functional activations for both the dual n-back task and a Raven’s matrices task in the same sample. Stated simply, do the patterns of activation observed for these tasks of WM and Gf support the notion that the former might be used to train the latter? The suggestion in the literature has been that if trained and untrained tasks individually and selectively activate overlapping brain regions, they ought to therefore engage overlapping psychological processes as well. In so doing, according to this logic, the training task (i.e. dual n-back) ought to exercise these same psychological processes in such a way that increased behavioural proficiency is conferred on the transfer task (i.e. RPM-type tasks, and Gf more generally; Dahlin, Neely, Larsson, Backman, & Nyberg, 2008; Buschkuehl, Hernandez-Garcia, Jaeggi, Bernard, & Jonides, 2014). Whether or not a single task ought to be taken to represent an entire domain of cognition and intelligence has been a matter of some debate, and spurred some researchers to propose a latent construct modelling approach (see Colom et al., 2013).
Buschkuehl and colleagues (2014) investigated the neural correlates of WM training with fMRI before and after participants trained with the n-back task for seven days. They found increased profusion in right postcentral gyrus, left superior frontal gyrus, as well as superior and middle occipital gyri bilaterally. Though their investigation did not include behavioural or fMRI measurements of RPM-type transfer tasks, they noted the similarity between their observed areas of increased activation following training and those known to be activated by RPM-type tasks.

They remarked: “…our results provide an excellent rationale for generalized cognitive improvement in visuospatial reasoning following WM training” (pp. 158). Our results largely support the overlap in neural activation between the dual n-back and RSPM tasks as predicted by Buschkuehl et al. (2014). Both the dual n-back task and the Raven’s Standard Progressive Matrices (RSPM) task both appeared to rely on key areas of the frontal and parietal cortices that have been described as the ‘core WM network’ (Owen, McMillan, Laird, & Bullmore, 2005), or the ‘central executive network’ (Rottschy et al., 2012), vital in a wide variety of cognitively demanding tasks. However, our results also provide reason for caution in further asserting that overlapping neural activations inexorably lead to training of constituent psychological processes.

First, as discussed above, neural activation reliably observed in the dual n-back task is quite similar to that observed in a wide range of cognitive tasks outside of the WM training literature. In this sense, what might be termed the ‘overlapping activation hypothesis’ of WM training outlined by Buschkuehl et al. (2014) becomes somewhat strained. That is, if overlapping activation on two tasks leads to a kind of psychological or cognitive exercise that increases performance on those tasks (one trained, one un-trained), then ought we to expect that any cognitively demanding task which activates this pattern of seemingly highly task-general cortical real-estate should also train and improve performance on tests of Gf? To be fair, this may in fact
be the case (and is indeed a testable hypothesis), however it certainly is not the claim of the present WM training literature. Rather, the claim is that training of WM specifically, is responsible for observed increases in $G_f$ after training. Additionally, the ‘overlapping activation hypothesis’ simply stated also allows for the unlikely possibility that training on $G_f$ tasks ought to conversely improve performance on tasks in the domain of WM (e.g. RPM-type tasks could train WM capacity). Again, this is a testable question, but not one that is currently of interest to the WM training literature, which is currently very much based on the overlapping psychological processes involved in WM and $G_f$ as outlined by Halford and colleagues (2007).

Given this critique, tasks which precipitate activation of similar cortical areas as the dual n-back task but are highly unrelated to WM or $G_f$ in terms of psychological processes should speak to the reasonability of the overlapping activation hypothesis. Our Lexical Decision task provides a fortuitous example in this regard. While classifying letter strings into words or non-words would not typically be considered a task of working memory or certainly $G_f$, our analyses revealed activation in areas quite similar to those of the dual n-back task and RSPM when task demands increased, including: bilateral medial frontal cortex, insula, anterior cingulate, basal ganglia, as well as left superior parietal areas. Despite this apparent overlap in neural activation, decades of cognitive psychology would suggest that these tasks are in fact representative of very discrepant domains of psychological functioning which ought not to influence one another in either direction via practice or repetition (i.e. crystallized intelligence versus fluid intelligence; Gottfredson & Saklofsky, 2009). Rather than to suggest that lexical decision making involves similar psychological processes to dual n-back or RSPM on the basis of this observed pattern of neural activation, a more likely explanation here is that all novel and difficult tasks robustly activate domain- and process-general areas of the brain. Recent work by Fedorenko, Duncan, and
Kanwisher (2013) has carefully identified a network which shows increased activation for hard > easy contrasts across a wide range of cognitively demanding tasks, including: precentral gyrus, anterior insula/frontal operculum, middle frontal gyrus, intraparietal sulcus, supplementary motor area, pre-supplementary motor area, anterior cingulate cortex, as well as posterior temporal and adjacent occipital regions. Interestingly, though perhaps not surprisingly, this primarily frontal and parietal network very closely resembles the ‘multiple-demand network’ discussed earlier, the network associated with Jung & Haier’s (2007) P-FIT model of intelligence, as well as the regions evoked by harder > easier contrasts for all three tasks fMRI tasks utilized in the present trial.

Furthermore, even when tasks involve similar neural activation patterns and psychological processes, the rationale for one having the capability to train the other still requires careful consideration. For example, in a meta-analysis of imaging studies in ‘deception’, Christ and colleagues (2009) found significant activation across studies in swaths of cortex highly consistent with the multiple-demand cortex activated by our dual n-back task. This is not wholly surprising, as the psychological processes involved in lying or deceiving others are likely similar to those involved in the n-back task itself: executive control/WM (e.g. keeping the truth in mind while conceptualizing deceptive responses), inhibitory control (e.g. supressing truthful responses), and task switching (e.g. alternating between truthful and un-truthful responses). Despite this however, it is unlikely that routinized practice in deceiving others should contribute to greater abilities in the domain of Gf due to overlapping patterns of neural activation. Thus, the overlapping activation hypothesis alone, unaided by knowledge of overlapping psychological or cognitive processes, would seem to have us make uncomfortable or otherwise unreasonable assertions about the kinds of tasks that ought to occasion increases in Gf by training of WM.
Stated differently, positing that a task can elicit behavioural increases in another on the basis of observed neural overlap, and presupposed psychological overlap, is simply a veiled instance of reverse inference - i.e. erroneously deducing psychological or cognitive functioning on the basis of observed neural activation (Poldrack, 2006). Importantly, none of these arguments about the tenability of the overlapping activation hypothesis are to say that WM training does not work, nor that its proposed mechanisms have no basis in specific neuroanatomy. Rather, that we should not expect WM training to generalize to behavioural increases on tasks of *Gf* on the basis of overlapping neural activation alone. If WM training does indeed work, to describe the mechanism(s) by which it does so as an overlapping of neural activations, simply will not suffice as an explanation.

So then, if WM training, or some other form of cognitive training is found to work for some population or other in future studies, by what mechanism might it work? The second main insight offered by our neuroimaging analyses is the reduction in load-dependant activation in the fronto-parietal network for the trained task (i.e. dual n-back) after WM training, but not for the task to which that heightened ability is purported to transfer to (i.e. RSPM task; Clark, Lawlor-Savage, & Goghari, in preparation). While our neuroimaging data helpfully point to where in the brain these training-related changes are happening, they are largely silent regarding the mechanisms of change at the cellular level (e.g. changes in the strength of synapses, changes in the number of dendritic spines, changes in the arborisation of axons, and changes in axon myelination; Constantinidis & Kingberg, 2016). Regardless, knowing that training-based task facilitation is reliably associated with fronto-parietal reductions in load-dependant activations allows us to speculate an important extension or modification to the overlapping activation hypothesis. In particular, we posit that if WM training were found to be effective in future, for
healthy young adults or other populations, the transfer task(s) of interest ought to show the same reductions in task-based fronto-parietal activation as the training task itself. It remains possible however, that such transfers of activation reduction will not be found for appreciably different tasks – even if superficially very similar. Reliable transfer of WM core network activation reductions may prove to be as elusive as far-transfer of cognitive ability itself.

It is tempting to additionally suggest here that fMRI investigations of fronto-parietal task-based activation patterns might be used as an exploratory paradigm in effectively matching training tasks to transfer tasks of interest. Neuroimaging might be effectively utilized in essentially developing more specific targets for cognitive intervention. However even if successful pairs of training and transfer tasks were to be discovered in this way through neuroimaging and neuroscience, it would still fall to the theory and practice of experimental psychology to draw conclusions about the affected cognitive domains as well as the ultimate educational, occupational, and/or social benefits of such facilitation of performance.

At root, improving cognition is an inherently behavioural endeavour, and must be evaluated primarily on the basis of its behavioural effects, through traditional methods of the behavioural sciences. In doing so, we cannot forget that the ultimate aim of improving intelligence is not to change synapses or axons, nor to gain points on psychometric tests of theoretical measures of one construct or another. Rather, through these latent and intermediary metrics, the ultimate aim of improving intelligence is that which is so succinctly summarized by Lumosity’s ‘deceptive’ advertising: “improve[ing] memory, attention, processing speed and problem solving skills so you can feel more confident in your abilities” (emphasis added; quoted in Redick et al., 2013; pp. 359). Improving cognition is only valuable to the extent that it is able
to benefit our behaviour, and our subsequent ruminations of it. William James (1881) expresses this most eloquently:

The current of life which runs in at our eyes or ears is meant to run out at our hands, feet, or lips. The only use of the thoughts it occasions while inside is to determine its direction to whichever of these organs shall, on the whole, under the circumstances actually present, act in the way most propitious to our welfare. …or in plainer English, perception and thinking are only there for behaviour’s sake. (reprinted in James, 1960, pp. 114)

The present WM literature fails this basic test. For all the theorizing about potential effects of WM training on mediating mechanisms or constructs (i.e. synapses, or myelination, or arborisation, or WM capacity), the observed effects fall short of reliably altering participants’ behaviour in ways ‘propitious to their welfare’ – placebo effects notwithstanding. In James’ terms, the current of WM training tasks that run in at participants eyes and ears do not appear to reliably or meaningfully run out at their hands, feet, or lips despite potentially causing some interesting phenomena amongst intermediate constructs whilst inside. Indeed, William James (who was amongst the first to reject the notion of meaningful transfer of cognitive abilities) would no doubt more satisfied with the claims posted on Lumosity’s website today versus those of 2013: “Enjoy brain training created by scientists and game designers” (Lumostiy.com).

Given the present and accumulated weight of evidence against reliable transfer effects of WM training, researchers should now abandon the idea that fundamental cognitive capacities in healthy young adults can be increased through repetitive engagement with computerized WM training tasks. Broadly speaking, people appear to get better at the particular tasks they train with, however, this increased ability appears to stubbornly and doggedly remain confined to that same task – which is rarely if ever the motivation behind such training. This is not to say that transfer of cognitive ability is impossible, but rather just that training of WM must unfortunately
be added to the long list of behavioural interventions and manipulations that do not produce this much sought after effect. Furthermore, even if WM training were to reliably increase performance on psychometric tests of fluid intelligence in methodologically rigorous experimental designs, it bears repeating that these tests remain an intermediate measure of the proposed true benefit of cognitive training (i.e. increased intelligence). This increase in psychometrically defined cognitive ability is in turn simply an estimated proxy for the kinds of educational, occupational, economic, and social outcomes that training would seek to augment by increasing general intelligence (e.g. school attainment, job success, and even longevity; Gottfredson, 1997; Gottfredson & Saklofske, 2009). Needless to say, repetitive computerized training of WM has not demonstrated benefit in these more distal and meaningful domains of functioning, having not yet passed the test of reliably benefitting the intermediary variables and constructs discussed above. Given this empirical repudiation of WM training effectiveness, future investigations in cognitive facilitation should instead focus on 1) alternative populations in which WM training may have reliable effects (see Ansari, 2015 for a review, and Weicker et al., 2016 for a meta-analysis); or 2) alternative interventions, or combinations of interventions which may help to produce the desired facilitative effects in healthy young adults (see McCabe, Redick, & Engle, 2016; Mrazek et al., 2016 for examples in this regard).
References


Thorndike, E. L., & Woodworth, R. S. (1901c). The influence of improvement in one mental function upon the efficiency of other functions III. Functions involving attention, observation and discrimination. *Psychological Review, 8*(6), 553-564. doi:10.1037/h0071363