Cognitive Neuroscience of Training and Transfer in Working Memory and Visual Attention

by

Todd Wesley Thompson

B. A. Rice University (1996)

Submitted to the Department of Brain and Cognitive Sciences in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy in Neuroscience

at the
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2015

Signature redacted

Signature of Author ....
Department of Brain and Cognitive Sciences
Signature redacted December 18, 2014

Certified by ....
John D. E. Gabreli
Grover Hermann Professor in Health Sciences and Technology and
Signature redacted Cognitive Neuroscience
Thesis Supervisor

Accepted by..
Matthew A. Wilson
Sherman Fairchild Professor of Neuroscience
Director of Graduate Education for Brain and Cognitive Science

Cognitive Neuroscience of Training and Transfer in Working Memory and Visual Attention

Todd Wesley Thompson

Submitted to the Department of Brain and Cognitive Sciences
on December 18, 2014 in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

ABSTRACT

The trained enhancement of working memory and visual attention has both theoretical implications for understanding the architectures of cognition, as well as practical implications for education and clinical treatment. In particular, transfer of training from one task to another may reveal shared psychological processes or neural systems across domains of cognition. In three experiments presented here, participants underwent a month of intensive training on either a complex working memory task or a visual attention task. Although participants made substantial gains on the trained tasks, that training did not yield transfer to untrained tasks measuring fluid intelligence, reading comprehension, or processing speed (Experiment 1). Brain imaging conducted before and after training revealed that increased working memory performance was accompanied by decreases of functional activation within anatomically circumscribed regions of frontal and parietal cortex as well as more wide-spread increases in frontoparietal functional connectivity (Experiment 2). Visual attention training using adaptively adjusted speeds on a multiple object tracking task revealed sizeable gains on the task itself, and those gains enabled the tracking of an increased number of items at a constant speed. This transfer from speed to quantity suggests that a common process underlies tracking speed and tracking capacity in visual attention (Experiment 3).

Thesis Supervisor: John D. E. Gabrieli
Title: Grover Hermann Professor in Health Sciences and Technology and Cognitive Neuroscience

Table of Contents

Acknowledgments

I would like to thank many people for their help and support throughout graduate school and, in particular, during the writing of this thesis. Firstly, I would like to thank my advisor, John Gabrieli, for his intellectual guidance, his broad curiosity, and especially for his seemingly boundless patience. Many thanks are due to the rest of committee, as well: to Rebecca Saxe, for her keen insight into seemingly intractable problems, for her friendship, for fascinating discussions about fMRI methodology and the purpose of science, and for being willing to adopt me into her lab; to George Alvarez, for maintaining his enthusiasm, friendship and humor while simultaneously suggesting yet another re-write of a chapter introduction; and to Nancy Kanwisher, for her thoughtfulness, wisdom, shared love of Echo, and her reliable ability to ask the important and hard questions underlying all of my research.

I would like to thank all of the people in all of the labs I've partially or wholly joined, for making MIT such an amazing place to get a PhD, in particular to SaxeLab, OlivaLab and GabLab. Special thanks are due to Amy Skerry, Tim Brady, and Jorie Koster-Hale, for their amazing encouragement and support writing this thesis. Without them and their last minute help, I suspect I would be still be writing. Thanks to my co-author on all of these manuscripts, Michael Waskom, for his frequent reminders, notices, and pointed interrogatories about the status of my thesis in person, over email, and across every messaging application in current use as of 2014. His prodigious programming skill made the analyses and figures in this research far better than they would have otherwise been.

Thanks, finally, to my family. To my birth family, for care packages and moral support, and to my chosen family, Jessica Paulus, Echolalia FiascoPants Thompson, and Claudine Petunia Porkchop Paulus. Jessica, you have the patience of a saint, are amazing in every way, and I am incredibly lucky to have found you. Echo and Claudine, you've taught me that with dogs, all things are possible. You are both very good dogs, and deserve all of the treats and long-awaited walks outside.

Chapter 1: Introduction

The challenges of training experiments

Typical experiments in cognitive neuroscience acquire data from each participant once and only once. One data point per subject is sufficient to answer most of our commonly asked questions: How does a patient group differ from a control group? Which brain regions does task A activate that task B does not? What is the relationship between a neural metric and a behavioral one? How are multiple behavioral metrics inter-related? Collecting one data point per subject is also convenient. Given a fixed amount of resources, it is possible to collect data on a broader sample of the population if each subject is only measured once, allowing broader inferences and, in many cases, more experimental power. Experiments with multiple sessions have very real problems with subject boredom and subsequent attrition, with finding appropriate testing materials with multiple forms so that test/retest effects are avoided, with more complicated counter-balancing requirements on experimental paradigms and, commonly, with the demands on experimenter and participant time.

Given all of these drawbacks, then, what are the reasons to invest in experiments involving prolonged training programs? The most compelling reasons are the types of questions that can be asked with training programs that cannot be asked with traditional experiments. Three general categories of question are more easily addressed with training studies, and each category is addressed by an experiment within this dissertation.

Outline of Dissertation

In experiment 1, we blend a scientific pursuit with a potentially therapeutic one. For decades, debate has raged about the relationship between executive functions, fluid intelligence and working memory, with some scientists claiming that fluid intelligence was a literally identical construct to working memory, simply with a different name. If this were true, we should expect that any training that produces improvements on a working memory task should yield comparable improvements on an untrained task measuring fluid intelligence. . In addition to the information gained about the relative independence of fluid intelligence and working memory, a training regimen capable of providing a measurable improvement on fluid intelligence (and the other real-world tasks correlated with it) could have monumental impact on life outside of academia, and suggests one way in which training experiments can provide a bridge between the basic sciences and the educational and rehabilitative fields. Using an intensive adaptive working memory training program, we test this hypothesis and discover that, in spite of a greatly expanded capacity on the trained task itself, no transfer is observed to any other measure of complex working memory, fluid intelligence, reading comprehension or processing speed measures.

Despite the lack of transfer in experiment 1, participants gained the ability to perform remarkably well at a dual n-back task, in some cases learning to successfully complete a dual 9-back. Using fMRI measurements collected before and after training, we report the changes in functional activation and in functional connectivity that underlie (and perhaps enable) the profound changes in task performance achieved with training, reported here as experiment 2.

In experiment 3, we use adaptive training multiple object tracking task in an effort to tease apart two competing theories of visual attention. In the attentional tracking literature, two limits on behavioral performance can be observed: the speed at which a fixed number of objects can be tracked and the number of objects that can be simultaneously tracked. Neuroimaging findings posit that these limits are largely independent, emerging from distinct neural substrates (with capacity supported by parietal cortex and speed supported by tracking processes in the frontal eye fields). In contrast, behavioral observations suggest that these two limits are in fact tightly coupled in an inverse relationship. By training one limit, we are able to determine if the other limit also improves, suggesting a shared resource. We find that training on speed does indeed yield transferred benefits to the number of items that can be tracked, supporting the shared-resource hypotheses of multiple object tracking.

Experimental Background

Experiment 1: Failure of Working Memory Training to Enhance Cognition or Intelligence

Historically, the academic study of intelligence and executive functioning (EF) has been plagued with difficulties. Some of these difficulties have been self-inflicted — the field's history is rampant with racist and eugenic overtones (Cattell, 1936), but other difficulties remain currently intractable. One particularly frustrating issue is that no one has been able to offer a working definition of fluid intelligence (Gf) or executive functioning that achieves academic consensus. Some conflate the two concepts entirely with working memory (WM), based on a robust correlation between performance on EF tests (e.g., the Towers of Hanoi, or Wisconsin Card Sorting tasks), performance on Gf tests (e.g., Raven's Progressive Matrices)(Blair, 2006;

McCabe, Roediger III, McDaniel, Balota, & Hambrick, 2010), and performance on "pure" working memory tasks (e.g., reverse span tasks). Other researchers argue for a more parcellated model based on latent factor analysis, usually separating working memory and inhibitory control from fluid intelligence, but still leaving both executive functioning and fluid intelligence essentially undefined (Heitz, Redick, & Hambrick, 2006; Miyake et al., 2000). Other researchers, still, argue that "intelligence" as a singular concept is wrong, and that there are actually multiple intelligences that are rarely, if ever, measured by psychometricians (Gardner, 1985). For these researchers, the connection between the multiple intelligences and executive functions remains, at best, murky.

Regardless of the difficulty establishing a precise academic definition, the broader importance of whatever EF and IQ tests are measuring has been well established outside the laboratory. Children who are able to delay eating a marshmallow in order to obtain additional marshmallows later score hundreds of points higher on their SATs (Mischel, Shoda, & Rodriguez, 1989). Additional IQ points correlate with additional years of life (Whalley & Deary, 2001), a greater income (Judge, Hurst, & Simon, 2009), and a decreased risk of a host of negative life outcomes (Gottfredson, 1997). In short, though we don't know exactly what we're measuring with our tasks, or even how many distinct cognitive concepts we're measuring, we appear to be measuring something important.

Given this real-world importance, one obvious question is whether executive function is susceptible to training. In other words, if an increased IQ yields dividends in the form of an increased salary, can we ameliorate poverty directly through a task that improves intelligence? In adults, the almost dogmatic answer to this question has been, "No." Whether intelligence is determined by nature, by nurture, or through some weighted combination of the two, by the time a child has become a teenager, their intelligence has been seen as fixed. As evidence for this point, it is noted that IQ tests administered in the teen years are almost perfectly correlated with IQ tests well into adulthood (Deary, Whalley, Lemmon, Crawford, & Starr, 2000).

More recently, a growing body of empirical work has claimed to observe improvements in EF under experimental conditions in healthy adults. Over the past decade, we've seen the field move from tentative suggestions about far transfer, with a focus on near transfer (e.g., from working memory tasks to working memory in other domains) to research claiming that adaptive training can improve the entirety of fluid intelligence. One of the earliest examples of training claiming to show transfer was in Klingberg's work, which attempted to alleviate ADHD symptoms through adaptive working memory training. In his experiment, a small number of children (n=7) diagnosed with ADHD underwent ~25 days of training on a set of adaptive working memory tasks (prominently featuring an adaptive spatial span task), with pre- and post-cognitive testing administered (Klingberg, Forssberg, & Westerberg, 2002). Compared to a control group who trained using a non-adaptive protocol, the researchers observed significant improvements in the trained version of the spatial span task and also in a physical variant of the span task. The more novel finding was that tests measuring other dimensions of cognitive control also improved. Accuracy on a Stroop task was significantly better in the trained group, while it remained unchanged in the control group. More generally, still, scores on the Raven's Progressive Matrices (RPM) improved selectively for the children who underwent adaptive training. Finally, in an attempt to measure the impact of training on hyperactive ADHD symptomology directly, subject's head-movements were measured during administration of a continuous performance task, and the group that received adaptive WM training showed significantly less movement after training, while the control group's movements were unchanged.

This initial study suffered from a very small sample size and a lack of blinding, leading to the possibility of experimenter expectancy effects driving the observed results. To address these concerns, the researchers performed a larger doubly-blinded study using the same training techniques and outcome measures. The main effects replicated— the treatment group showed improvements on the trained computerized spatial span task, on the physical implementation of that task (near transfer), and again on the Stroop and Raven's Progressive Matrices (far transfer) (Klingberg et al., 2005).

Klingberg modified and improved his tasks from the rudimentary laboratory prototypes into a commercial training tool, Cogmed, that incorporated new adaptive training tools across other working memory modalities, improved graphics and feedback screens, and other bells and whistles. Using this program, Holmes and colleagues attempted to use cognitive training as a way of ameliorating the working memory deficits commonly afflicting those who perform poorly in academic settings (Holmes, Gathercole, & Dunning, 2009). After twenty days of training using the Cogmed program, 10-year old children improved on measures of standardized WM to within normal ranges, while children in a non-adaptive control group did not significantly improve. Furthermore, these gains were maintained at a six-month followup.

Unlike the previous experiments, however, the participants in this experiment did not show immediate "far" transfer to measures of verbal or performance IQ, though the authors note that there were significant improvements in mathematical reasoning (presumably a task loading on fluid intelligence) at the six-month followup. They maintain that the boosts in underlying working memory abilities enabled the children to progress more efficiently in the academic environment, and thus gain previously missing mathematical skills. While this does account for the performance lag in the mathematical reasoning outcome measure, it remains unclear how to reconcile the immediate lack of far transfer with the other studies that report finding far transfer.

Jaeggi, et al., finally brought the hope and promise of cognitive training into the mainstream psychological literature. Using an adaptive dual n-back task as their training tool, they explicitly asked whether training working memory could show far transfer to unpracticed measures of fluid intelligence (Jaeggi, Buschkuehl, Jonides, & Perrig, 2008), measured (as is typical) by matrix reasoning tests. In adults, they show that twenty days of training on the n-back task not only shows the now-expected improvement on the task itself, but also confers marked improvements in the fluid reasoning outcome measures. Furthermore, the benefits on the fluid reasoning task accrue in a dose-dependent fashion — the improvements seen with 3 days of training are far smaller than those after a week of training, and the week's gain on the matrix problems pales in comparison to the gain observed after twenty days of training.

While this paper received a huge amount of publicity and enthusiasm, some qualifications were raised about Jaeggi's methodology. Moody, in particular, challenged their fluid reasoning outcome measure (Moody, 2009). Typically, matrix reasoning tasks are administered without intense time pressure. The RAPM is designed to be completed in 45 minutes, for instance. In the Jaeggi study, participants only had 10 minutes to complete as many problems as possible before and after the training. This deviation from standard testing protocol raises a concern that the matrix reasoning task is no longer a "pure" measure of fluid intelligence, but is now more heavily weighted by speed of processing or a shift in the speed/accuracy strategies. However, based on unpublished data, Jaeggi argues that the link between her abbreviated implementation of the matrix tasks and the standard implementation is robust, and has continued to use the abbreviated task as a primary outcome measure.

The controversies surrounding the transfer from working memory to measurements of fluid intelligence continue unabated. Although Jaeggi reports additional replications of transfer from working memory training to fluid intelligence (see [Au et al., 2014] for a recent meta-analysis), our own results from experiment 1 demonstrate that that transfer is far less robust than would be hoped if working memory training is to make a substantial impact outside of academia.

Experiment 2: Neural Correlates of an Intensive Working Memory Training Program

Perhaps the research questions training paradigms are most obviously capable of addressing are those questions exploring the training itself. A rich literature exists exploring the neural correlates of training motor and perceptual skills (see Kelly & Garavan, 2005 for a review). Typical findings support the "scaffolding-storage" model proposed by (Petersen, van Mier, Fiez, & Raichle, 1998), in which domain-general attentional and control regions (including prefrontal cortex, anterior cingulate cortex and posterior parietal cortex) support learning of a new, effortful task. This "scaffold" then falls away as domain-specific sensory and motor "storage" regions develop more refined task representation and become able to support task performance on their own. This "scaffold-storage" model makes clear predictions when the trained task is supported by domain-specific sensory or motor areas, but leaves an interesting question unaddressed: what happens during prolonged training of domain-general processes (e.g., working memory)? In this case, when the domain-general regions continue to be responsible for task performance even after training, will the pattern of reduced activation persist?

In general, neuroimaging results from training experiments report one of four patterns of training-related change (Buschkuehl, Jaeggi, & Jonides, 2012; Kelly & Garavan, 2005):

1. Increased activation — in the regions supporting the trained function, more activity is observed, usually interpreted as a greater capacity for neuronal recruitment as the task is practiced.
2. Decreased activation — there is reduced activity in the regions supporting the trained function, possibly interpreted as greater "efficiency" at performing the task
3. Activation redistribution — Both of the above patterns are seen simultaneously. If a task activates, for example, both prefrontal and motor cortices, we might see the domain-independent prefrontal activation decrease with practice while the purportedly more task-specific motor activations increase, or,
4. Reorganization of networks — With practice, fundamentally different brain regions are recruited to support the task.

In theory, working memory training could show any of these patterns. Unfortunately, different working memory training studies appear to support different hypotheses. Garavan and colleagues (Garavan, Kelley, Rosen, Rao, & Stein, 2000) document a decrease in 14 out of 17 clusters activated by a delayed match-to-sample task within four hours of practice. This pattern of activation decrease within a single scanning session was subsequently replicated by Jansma, et al. (Jansma, Ramsey, Slagter, & Kahn, 2001), Landau, et al. (Landau, Garavan, Schumacher, &

D'Esposito, 2007; Landau, Schumacher, Garavan, Druzgal, & D'Esposito, 2004), and Sayala, et al. (Sayala, Sala, & Courtney, 2006).

Klingberg (Klingberg, 2010) noted that all of these experiments examined only the earliest stages of working memory training, and suggested that the observed activation decreases were a signature of the short timeframe, in which changes reflecting strategy learning and task familiarity would dominate any changes accompanying genuine increases in working memory capacity. In contrast, longer WM training programs (Dahlin et al., 2008; Jolles, Grol, Van Buchem, Rombouts, & Crone, 2010; Olesen, Westerberg, & Klingberg, 2004) all reported increased activity in task-relevant regions, supporting Klingberg's hypothesis that, unlike motor and perceptual learning, training on working memory paradigms would result in persistent increased neural activity accompanying the increased cognitive capacities. More recent WM training work (Schneiders, Opitz, Krick, & Mecklinger, 2011; Schweizer, Grahn, Hampshire, Mobbs, & Dalgleish, 2013), however, reports a different finding – *decreased* activations in domain-general prefrontal and parietal regions, refuting Klingberg's hypothesis of strictly increasing activation after WM training.

It is becoming clear that sweeping and generic statements about the neural outcomes of WM training are unlikely to be accurate, and that, instead, changes in functional activation are likely to be specific both to the type of training and, perhaps, to the pre- and post-training performance. In the case where a task can not be reliably completed pre-training but is possible after training (as in [Jolles et al., 2010; Olesen et al., 2004]), an activation increase may simply reflect an artifact of time-on-task. Alternately, the relationship between the time of imaging and the trajectory of task acquisition may influence the results. Imaging during a single-session may exhibit functional decreases reflecting comfort in the scanning environment, a reduction in the salience of a novel task and the mastery of ancillary task demands (e.g., mapping between responses and button-box keys). Scanning during the middle of a longer-term adaptive training experiment could reveal increased activity enabling increased performance, while scanning at a long-mastered difficulty level after training may show activation decreases relating to what is commonly called increased "neural efficiency" (Richard J. Haier, Siegel, Tang, Abel, & Buchsbaum, 1992).

In order to address some of these confusions, in Experiment 2 we acquired imaging data from participants who had already practiced a dual n-back task (Jaeggi et al., 2008) outside of the scanner, and focus our analyses on difficulty levels that are accessible (but difficult) before training and become simple after training.

Experiment 3: Training Multiple Object Tracking Speeds Improves Capacity Limits

There are many paradigms with which to study visual attention, but one of the most commonly used is multiple object tracking (MOT) (Pylyshyn & Storm, 1988). In these experiments, participants are presented with a display of identical objects. Some of those objects are identified as targets to be subsequently tracked, after which the entire set of objects begins moving. When the movement stops, the participant is asked to identify the original target objects. As Scholl (Scholl, 2009) describes, this task has several key features important for the study of attention: 1) it requires sustained attention, lest the target items be lost in a moment of distraction; 2) it requires attention to multiple objects at the same time, allowing experimenters to distinguish between models of visual attention incorporating a single unitary "spotlight" and those allowing for multiple spotlights; 3) it has intuitive ecological validity to everyday human attention, where it is commonly required to monitor several moving objects in order to, say, navigate a busy mall parking lot.

One quite straightforward question that can be asked with this paradigm is whether the capacities it measures are amenable to improvement at all. Other measurements of visual attention such as change blindness have proven quite resistant to improvement on untrained stimuli (Gaspar, Neider, Simons, McCarley, & Kramer, 2013), while others are more tractable (e.g., perception of Gabor orientation, (Lu, Chu, Dosher, & Lee, 2005)). Can performance on a multiple object tracking task be improved at all?

A more theoretically interesting question is whether training can be used to discriminate between two plausible theories of visual attention. In the MOT task, two limits on visual attention can be extracted. The first, a *speed limit*, describes the speed at which a fixed number of targets (usually four) can be successfully tracked. The second, a *number limit*, describes the number of target items that can be tracked at a fixed speed. Several behavioral experiments suggest that speed limits and number limits are not independent: the speed limit on tracking depends on the number of targets, and number limit depends on their speed (Alvarez & Franconeri, 2007; Holcombe, Chen, & Howe, 2014). A different theory emerges from the neuroimaging literature, however, which suggests that speed limits and number limits may ultimately be rooted in separate underlying neural mechanisms (Culham, Cavanagh, & Kanwisher, 2001; Shim, Alvarez, Vickery, & Jiang, 2010). More specifically, activation levels in the posterior parietal cortices reflect changes in quantity while remaining insensitive to changes in speed. The frontal eye fields, in comparison, *are* sensitive to speed. One implication from this finding is that the speed limits and capacity limits may reflect separate and independent limits on attentional tracking.

While these two competing hypotheses are challenging to resolve through traditional neuroimaging or individual differences experiments, they make specific predictions about the result of a training study that improves tracking speed. If the number of tracked items is limited by attentional pointers residing in the speed-insensitive parietal cortices, as suggested by (Culham et al., 2001), speed-training should not increase the number of tracked items. On the other hand, if speed limits and number limits represent two facets of a common resource, successfully increasing the speed at which items can be tracked should directly improve the number of items that can be tracked at a fixed speed.

Chapter 2: Failure of Working Memory Training to Enhance Cognition or Intelligence

Abstract

Fluid intelligence is important for successful functioning in the modern world, but much evidence suggests that fluid intelligence is largely immutable after childhood. Recently, however, researchers have reported gains in fluid intelligence after multiple sessions of adaptive working memory training in adults. The current study attempted to replicate and expand those results by administering a broad assessment of cognitive abilities and personality traits to young adults who underwent 20 sessions of an adaptive dual n-back working memory training program and comparing their post-training performance on those tests to a matched set of young adults who underwent 20 sessions of an adaptive attentional tracking program. Pre- and post-training measurements of fluid intelligence, standardized intelligence tests, speed of processing, reading skills, and other tests of working memory were assessed. Both training groups exhibited substantial and specific improvements on the trained tasks that persisted for at least 6 months post-training, but no transfer of improvement was observed to any of the non-trained measurements when compared to a third untrained group serving as a passive control. These findings fail to support the idea that adaptive working memory training in healthy young adults enhances working memory capacity in non-trained tasks, fluid intelligence, or other measures of cognitive abilities.

Introduction

A fundamental question of both theoretical and practical interest is whether the basic human cognitive abilities that underlie many aspects of learning, memory, thinking, and performance can be enhanced in adults. It has long been thought that the combination of genetics and early environment substantially determines life-long individual differences in generalizable cognitive abilities (i.e., abilities that support and limit performance on a wide range of tasks). Because standardized intelligence quotient (IQ) scores predict performance on a wide range of cognitive tasks and educational achievements (Deary, Strand, Smith, & Fernandes, 2007), IQ scores are often used as an index of general cognitive abilities. Such IQ measures exhibit substantial correlations from late childhood through adulthood (e.g., IQ scores were estimated to correlate 0.73 from ages 11 through 77 in a longitudinal study (Deary et al., 2000)). These observations suggest that variation in general cognitive abilities is determined, to a large extent, by late childhood or early adolescence. This fixedness of cognitive ability has seemed especially strong for fluid intelligence (the ability to solve novel problems), relative to crystallized intelligence (the ability to apply specific knowledge, skills, and experience). In part this is because scores on tests of crystallized intelligence can be improved by, for example, instructing a student on the vocabulary that the crystallized intelligence tests typically evaluate, but also in part because fluid intelligence has typically been considered as more biologically determined than crystallized intelligence (Neisser et al., 1996; Nisbett et al., 2012).

More recently, evidence has emerged indicating some plasticity in IQ and its neural bases. One study reported that verbal and performance IQ scores, as well as their neural correlates, exhibited some fluctuation across the teenage years, rather than remaining static (Ramsden et al., 2011). A particularly influential study by Jaeggi and colleagues not only reported plasticity in adult fluid intelligence, but also defined a specific cognitive training program that enhanced fluid intelligence (Jaeggi et al., 2008). In this study, young adults performed a working memory (WM) task for about 25 minutes per day for up to 19 days. The WM task trained WM capacity, defined here as the amount of goal-relevant information that could be simultaneously maintained and processed. Specifically, the training task used a "dual n-back" paradigm in which participants simultaneously heard letters and saw spatial locations presented one after another. Their task was to respond whenever a presented stimulus was identical to the stimulus presented *n* trials ago (e.g., in a dual 2-back, subjects responded whenever the current spatial position *or* the current auditory stimulus matched the presentation from 2 trials earlier). Performance improved on the trained WM task, and most importantly, there were significant post-training gains on a measure of fluid intelligence. Thus, the learned skill in performing the WM task *transferred* to a growth in fluid intelligence. These findings were exciting because they offered a way to enhance adult fluid intelligence, previously viewed as static. Because superior fluid intelligence is associated with superior performance on many cognitive and learning measures, these findings suggested a practical way by which cognitive training might lead to widespread gains in cognitive ability.

Two aspects of the WM training that yielded a gain in fluid intelligence seem important. First, it trained a cognitive construct (working memory) that has been associated with fluid intelligence in many studies (R. W. Engle, 2002; Kyllonen & Christal, 1990), such that transfer might be expected. Generally, transfer might be expected from one task to another when those two tasks share common cognitive mechanisms, either through reliance on similar cognitive processes, or through a shared neural substrate. Among adults, greater WM capacity is associated with superior performance in a broad range of high-level cognitive domains, including reading comprehension, problem solving, and inhibitory control (Andrew R.A. Conway, Kane, & Engle, 2003) and so is thought to reflect central executive capability (R. W. Engle, 2002). Thus, it is plausible that WM training might improve central executive capability and/or fluid intelligence. Second, the WM training was adaptive, such that the span (or the number of intervening stimuli) increased between the presented target and its potential match as a participant performed better on the task, or decreased as the participant performed worse on the task. Such an adaptive design makes certain that the participant constantly performs at a challenging but not frustrating level. These types of adaptive designs have been a core feature of effective WM training (reviewed in (Klingberg, 2010)). Indeed, this adaptive design resulted in more than a doubling of WM capacity on the trained WM task (Jaeggi et al., 2008). Thus, the training program that raised fluid intelligence was theoretically motivated and effective in design.

The provocative finding that a WM capacity training task can increase fluid IQ in adults raised several questions (Sternberg, 2008). First, the control group was a no-contact group that was tested on the fluid IQ measure with a comparable testing interval. The lack of an active training regime for the control group leaves open questions of specificity (e.g., would any demanding training program yield such a gain in fluid IQ? are there correlated factors such as motivation associated with the training experience that influence transfer?). Second, transfer was only demonstrated on one specific test of fluid IQ, leaving open the question of the scope and limits of the transfer of cognitive gains from the WM training program (e.g., would such transfer occur for another measure of fluid IQ? would it occur for measures of crystallized IQ or other cognitive abilities such as processing speed?). Third, does such WM training result in enduring gains that are sustained well after the training program, or must the training be continued to maintain gains on either WM or fluid intelligence measures?

After publication of the Jaeggi et al. study (Jaeggi et al., 2008), several subsequent studies have examined the influence of WM training on fluid IQ and other types of cognition. One study, using a similarly adaptive WM training program, reported no gains on fluid IQ, but did report gains in reading and cognitive control (Chein & Morrison, 2010). Two other studies, using dual n-back training tasks identical to Jaeggi et al. (Jaeggi et al., 2008) failed to find any gains on fluid IQ (Chooi & Thompson, 2012; Redick et al., 2012). Other research was more consistent with the original findings, including (1) a partial replication in children, in which participants who exhibited gains on the WM training task also exhibited gains on a fluid IQ measure (Jaeggi, Buschkuehl, Jonides, & Shah, 2011); (2) a report of both fluid intelligence improvements and corresponding changes in EEG measures after WM training which included the dual n-back among other tasks (Jaušovec & Jaušovec, 2012); and (3) a finding of transfer from both single n-back and dual n-back training to fluid intelligence gains, but with effects mediated by conscientiousness and neuroticism personality factors ((Studer-Luethi, Jaeggi, Buschkuehl, & Perrig, 2012), originally reported in (Jaeggi et al., 2010)).

Because the transfer from WM training to fluid intelligence is both controversial and important, we aimed to replicate and extend the finding that WM training enhances fluid IQ. Two groups of young adults, stratified so as to be equated on initial fluid IQ scores, were randomly assigned to two conditions (a randomized controlled trial or RCT). The experimental group performed the dual n-back task (as in the original Jaeggi et al., 2008 study (Jaeggi et al., 2008)) for approximately 40 minutes per day, 5 days per week for 4 weeks (20 sessions of 30 blocks per session, exceeding the maximum of 19 sessions of 20 blocks per day in the original Jaeggi et al., 2008 study). An active control group performed a visuospatial skill learning task, multiple object tracking (or MOT), on an identical training schedule. We also tested a no-contact group equated for initial fluid IQ in case both kinds of training enhanced cognitive abilities.

Tests of cognition were administered before and after training (or after an equal duration of time for the no-contact group) in order to evaluate the benefits of the training. Two tests were versions of the training tasks (dual n-back and MOT). We hypothesized that, as in prior studies, there would be significant improvements on the trained tasks, and that because the tasks were quite different, there would be selective gains on the trained relative to the untrained tasks for both groups. We also asked in a subset of participants whether the skills gained during training would endure over a 6-month period without further training.

A second set of tests measured *near transfer*, gains on untrained WM capacity measures that were conceptually similar to the dual n-back training task. In Baddeley and Hitch's original model of working memory (Baddeley & Hitch, 1974), working memory has separate and independent slave subsystems (the *phonological loop* and *visuospatial sketchpad*), and these modality-specific storage systems are coordinated by a modality-independent *central executive*. Evidence for transfer from trained WM tasks to non-trained WM tasks suggests that these WM tasks share underlying processes (e.g., (Anguera et al., 2012; Brehmer, Westerberg, & Bäckman, 2012; Buschkuehl et al., 2008)). In the present study, we selected two widely studied tasks, Operation Span and Reading Span (Andrew R A Conway et al., 2005), which are similar to the dual n-back task because all three tasks measure complex working memory (CWM). All three of these CWM tasks involve encoding a presented stimulus, performing some sort of updating/manipulation (validating a math problem, assessing the sensibility of a sentence, or updating the numerical position of the rehearsed stimuli), and retrieval (either of all the encoded stimuli in the case of the span tasks, or of the nth-back stimuli in the dual n-back task). Transfer of any broad gain in WM capacity would be expected on the Operation Span and Reading Span tasks if dual n-back training enhances either the capacities of either the phonological loop (responsible for the storage of verbally encoded material for subsequent retrieval) or of the central executive (responsible for the updating and manipulation components of the tasks).

The Operation and Reading Span tasks were selected specifically because there is considerable evidence that these tasks measure the central executive component of WM. Performance on these tasks has been correlated with performance on a broad range of other tasks, including tests of verbal, numerical and spatial reasoning, matrix reasoning such as the Raven's Progressive Matrices, processing speed, and general knowledge (Ackerman, Beier, & Boyle, 2005; Kane, Hambrick, & Conway, 2005). Observing an improvement on these CWM measures following dual n-back training could lend support to the idea that dual n-back training increases CWM capacity.

In addition to the assessment of trained tasks and the near-transfer tasks, a third set of tests measured *far transfer*, gains on measures that were dissimilar to the WM training task, including measures of fluid IQ, crystallized IQ, reading skill, and processing speed. Although the common components between the dual n-back task and the far-transfer tasks are not as apparent as those in the near-transfer tasks, there are often strong correlations between measures of CWM and fluid intelligence, which suggests that there are shared mental processes (Andrew R.A. Conway et al., 2003; R W Engle, Kane, & Tuholski, 1999; Kyllonen & Christal, 1990). The prior report that training on the dual n-back task enhanced scores on matrix reasoning tasks further supports the idea that CWM capacity and fluid intelligence share underlying processes (Jaeggi et al., 2008). Additional measures of far transfer were selected to determine the scope and limits of transfer from WM training, as well as a specific report that similar training enhanced reading skills (Chein & Morrison, 2010).

We also examined the possibility of individual personality differences among participants modulating either training or transfer, in an attempt to illuminate the reasons behind the mixed results so far reported in the WM training literature. Greater conscientiousness has been reported to predict greater improvement on a dual n-back task during training, but lesser transfer of training to a measure of fluid intelligence transfer (Studer-Luethi et al., 2012). We therefore measured conscientiousness in all participants as the "Conscientiousness" factor from the Big Five personality test (Costa & McCrae, 1992). We also examined two additional characteristics of all participants. We measured implicit theories of intelligence, defined as the extent to which a person believes that intelligence is a fixed or innate trait, as opposed to viewing intelligence as a capacity that can incrementally grow through effort and learning. Those who view intelligence as improvable with effort are said to have a "growth mindset" (C. Dweck, 2006). We also measured "grit", defined as perseverance and passion for long-term goals (Angela L. Duckworth, Peterson, Matthews, & Kelly, 2007). Both growth mindset (Blackwell, Trzesniewski, & Dweck, 2007) and greater grit (Angela L. Duckworth et al., 2007) have been associated with better performance and learning in a variety of settings.

Methods

Participants, Recruitment, and Group Assignment

Participants were recruited through web advertisements, physical flyers, and e-mail to the Northeastern and Tufts college mailing lists. Participants were required to be adults between the ages of 18 and 45, right-handed, in good health, and not taking any drugs. All participants provided informed, written consent before participation. This study was approved by the Massachusetts Institute of Technology Institutional Review Board (PI: Leigh Firn).

After recruiting each participant, we performed pre-training behavioral testing and determined his or her group assignment (Table 1). Each incoming participant was paired with another participant based on age, gender, and score on the Raven's Advanced Progressive Matrices (RAPM) task, and each member of that pair was randomly assigned to either the n-back or the MOT training group. The No-Contact group was recruited separately, but in the same fashion, and matched to a training pair by gender and initial RAPM. Because of this matching procedure, the No-Contract group was slightly, but significantly, older than the two training groups (Table 1). The No-Contact group averaged 1.8 years older than the other two groups [$F(2,55) = 3.37$, $p < .05$]. However, the three groups did not differ significantly by gender or RAPM scores [$F(2,55) < 1$, $p > .8$], nor did they differ on the full IQ score from the Wechsler Abbreviated Scale of Intelligence (Wechsler, 1999), administered as part of the pre-training battery [$F(2,55) < 1$, $p > .4$].

Table 1. Participant Characteristics

Training Group	Average Age	Gender	RAPM (SD)	Full-4 IQ (SD)
Dual n-back	21.2	7 M, 13 F	13.3 (2.1)	120.8 (10.8)
Multiple Object Tracking	21.3	8 M, 11 F	13.6 (2.0)	120.7 (7.0)
No Contact	23.1	7 M, 12 F	13.3 (2.2)	117.6 (7.4)

Table 1. Participant Characteristics. Participants were assigned to treatment groups based primarily on gender and initial score (out of 17) on the Ravens Advanced Progressive Matrices problems (RAPM)

Eighteen potential participants either dropped out of the study or were excluded after initial testing was completed. Two participants assigned to the dual n-back condition voluntarily withdrew (one after 5 days of training, the other after 9 days); no other participants had begun training when they were excluded or withdrew. Five participants provided initial behavioral data during the process of collecting the passive-control group, but were not included because they were not well-matched to an unmatched member of the other two groups based on Ravens score. The remaining eleven subjects were not included for a variety of logistical reasons, including difficulties aligning schedules with the experimenters, claustrophobia or excessive movement in fMRI scanning sessions, or repeatedly skipping appointments. Although we attempted to perform all behavioral measures with all included participants, in a few cases there were technical problems in administering some measures to some participants (these are noted in Tables 2 and 3).

Participant payment

Participants in the training groups were paid $20 per training session, with a $20 bonus per week for completing all five training sessions in that week. All participants were paid $20 per hour for behavioral testing, and $30 per hour for imaging sessions (data from imaging sessions are reported separately).

Overall experiment design

After recruitment, participants underwent approximately six hours of behavioral testing spread across three days and two hours of structural and functional magnetic resonance imaging. If a participant was assigned to one of the two active training conditions, they then completed twenty sessions of adaptive training on campus.

After training was completed, post-training behavioral testing and imaging were administered as soon as possible. (Average number of days between last training session and post-training testing was 4.3 days, with a minimum of 0 days and maximum of 14 days. Two participants were tested on the final training day, with at least 3 hours between the last training session and the post-testing session; all other participants were tested at least a day after the last training session. This time was not significantly different between groups [$t(37) = .2$, $p > .8$]). Participants in the active training conditions were asked to return approximately six months after the completion of training later to examine the status of their improvement on the trained tasks. (Average number of days before the follow-up testing was 187 days, with a minimum of 122 days and maximum of 252 days. This time was not significantly different between groups [$t(19) = .78$, $p > .4$]). Although some participants in each training group were unable to return for follow-up testing (primarily due to post-graduation dispersal), data from 11 participants in the MOT training group and 10 participants in the n-back training group were collected. For behavioral measurements in which the participant's score was evaluated by the tester (e.g., the vocabulary sections of the WASI), testers were blinded to the participant's training condition.

Behavioral Testing - Trained Tasks

To establish baseline measures of the two possible training tasks and test for transfer or practice effects from one training condition to another, performance on both training tasks was evaluated before and after the training period.

Baseline Dual n-back -- Implementation of the adaptive dual n-back training task followed Jaeggi et al., 2008 (Jaeggi et al., 2008). An auditory letter and a visual square were simultaneously presented for 500ms, followed by a 2500ms response period. Letters were chosen from the consonants B, F, H, J, M, Q, R, and W to maximize auditory discriminability between letters. Squares were presented at one of eight positions evenly spaced around the periphery of the screen. Participants responded when one or both of the current stimuli matched a stimulus presented n trials ago. Auditory matches were identified with the index finger of the right hand, and visual matches were identified with the middle finger of the right hand. No response was required on trials that did not match the target, and either response could be made on trials where both stimuli matched. Each block presented $n + 20$ trials, containing four auditory target trials, four visual target trials, and two trials where both auditory and visual stimuli matched. For baseline testing of the dual n-back task, participants completed 30 blocks of dual n-back trials, with 5 blocks of each level from 1-Back to 6-Back presented in a counter-balanced pseudorandom order. Participants were allowed to take breaks between blocks as needed. In order to control for response biases between subjects, a sensitivity index (d') was calculated at each level from 1-back to 6-back for each participant (Wickens, 2001). Because participants in all three groups scored highly on the first level of the n-back without any practice, the dependent measure used to evaluate improvement was calculated by averaging the d' scores from 2-back to 6-back for each participant.

Multiple Object Tracking -- To assess the maximum speed at which participants could reliably track moving objects, we followed the general techniques from Alvarez & Franconeri, 2007 (Alvarez & Franconeri, 2007). Participants were asked to track 4 dots among 12 distractor dots. At the beginning of each trial, 4 target dots were identified in green for 500ms while all dots remained stationary. For the next 2500ms, all 16 dots moved while the target dots remained identified in green. At that point, the 4 target dots turned black, and for the remaining 8500ms of the trial, the target dots appeared identical to the distractor dots while the participant attempted to remember which dots were targets. Finally, participants identified the 4 tracked dots using a mouse and were given feedback.

The initial speed at which items moved for each participant was determined by a self-assessment task in which participants used the cursor keys to make targets move slower or faster and reported the speed at which they thought they could reliably track four targets. This was followed by a thresholding procedure over the following 90 trials in which the speed of the moving dots increased by .5 degrees of visual angle/second every time two trials in a row were answered correctly, and decreased by .5 degrees/second every time two trials in a row were answered incorrectly. To count as a correct trial, all 4 targets were required to be identified correctly. Participants were allowed to take breaks, as needed. The speed of the final trial was the dependent variable.

Near Transfer Tasks - Working Memory Capacity

Automated Operation Span (Complex WM Capacity) (Unsworth, Heitz, Schrock, & Engle, 2005) -- Participants were presented with alternating letters and math equations, and asked to remember the letters while assessing whether each math equation was valid. Set sizes ranged from 3-letters to 7-letters, with each set size presented for 3 trials over the course of the task, in a random order. At the end of each trial, participants reported the letters in the order they were presented. The dependent measure was the "score" variable reported from the ePrime program, which is the sum of all perfectly remembered letter sets. One dual n-back participant's pre-training Operation Span score was excluded for falling more than 3 standard deviations below the group average, whereas all of this participant's other behavioral measurements were near the group average, including the cognitively similar Reading Span score. It is unclear whether this score represented some sort of experimenter error in data collection or participant confusion about the task instructions.

Automated Reading Span (Complex WM Capacity) (Unsworth et al., 2005) -- Participants were presented with alternating letters and sentences, and asked to remember the letters while assessing whether each sentence was sensical. Set sizes and scoring were identical to the Automated Operation Span.

Combined Span Task (Complex WM Capacity) -- Scores from the Automated Operation Span and Automated Reading Span were summed to create a single measure estimating a participant's complex working memory capacity.

Far Transfer — Standardized Intelligence Tasks

Raven's Advanced Progressive Matrices (Fluid Intelligence) (Raven, JC, Court, JH, &

Raven, J, 1998) -- Each item presented a three-by-three grid filled with patterns, with the bottom-right entry missing. Participants selected the best of 8 choices to fill the missing location based on the pattern of the other elements in the matrix. We created two forms of the 36-item RAPM test for pre- and post-testing. The two forms were equated for difficulty based primarily on published accuracy rates per item (Raven, JC et al., 1998) and secondarily on pilot experiments assessing the average response time per item. Because the last two items of the test are much more difficult than the rest of the test and are not matched to each other in difficulty, the dependent variable was the number of correct responses out of the first 17 items. Form A consisted of items 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 22, 23, 25, 28, 29, 31, 34, and 36, while form B consisted of items 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 21, 24, 26, 27, 30, 32, 33, and 35. Participants were given 25 minutes to complete each half of the RAPM.

Wechsler Abbreviated Scale of Intelligence (WASI) /Wechsler Adult Intelligence Score III (WAIS-III) (Wechsler, 1999, 2005) -- The WASI and WAIS are commonly used assessments of standardized intelligence. Because they have been normed against each other and use common sub-tests with different forms, the two tests provide a simple way of acquiring a matched IQ measurement before and after training. Rather than counter-balancing the two tests, the WASI was administered pre-training and the WAIS post-training so as to maximize the sensitivity of measuring any training-related transfer between groups.

WASI/WAIS Blocks -- Participants were given a set of physical blocks with red and white shading on them, and asked to assemble them so as to replicate a target pattern. The amount of time needed to replicate the patterns was the raw score that was converted into a scaled score, which was then used as the dependent measurement.

WASI/WAIS Matrices -- Participants selected the best-fitting item to complete a grid of figures, based on abstract rules and relations between the other figures in the grid.

WASI/WAIS Vocabulary -- Participants were required to verbally define progressively more challenging vocabulary words.

WASI/WAIS Similarities – Participants were asked to relate pairs of concepts (e.g., How are a snake and an alligator alike?).

Far Transfer — Reading Comprehension

Nelson Denny Comprehension Subtest (Brown, Fishco, & Hanna, 1993) – Participants were asked to read five short passages and respond to 38 short questions about the contents of those passages.

Nelson Denny Reading Rate (Brown et al., 1993)– During the first passage in the comprehension subtest, participants' reading rate was assessed by recording the number of words read in the first minute.

Far Transfer — Speed of Processing

WAIS-III Digit/Symbol Coding (Wechsler, 2005) -- Participants were provided with a set of digit-symbol pairs and a list of digits. Under each digit, participants wrote down as many corresponding symbols as possible during a two-minute span.

Woodcock-Johnson III Tests of Cognitive Abilities: Visual Matching (Woodcock, McGrew, & Mather, 2001) -- For 3 minutes, participants scanned rows of numbers and circled the two identical numbers in that row.

Woodcock-Johnson III Tests of Cognitive Abilities: Pair Cancellation (Woodcock et al., 2001) -- For 3 minutes, participants scanned rows of figures and circled each instance in which a target picture was followed immediately by a second target picture (e.g., a cat followed by a tree.)

Personality measurements

Dweck Intelligence Questionnaire (C. S. Dweck, 2000) – Participants were asked to indicate the extent that they agree/disagree with 8 statements regarding the malleability of intelligence (e.g., "You have a certain amount of intelligence, and you really can't do much to change it") on a 5-point scale. The dependent measure is the sum of their answers (with some items reversed in scoring), with a lower score indicating a more static view of intelligence.

Conscientiousness Factor Questionnaire (Costa & McCrae, 1992) – Participants were asked to rate how well-described they were by 12 statements assessing their perception of their own conscientiousness on a 5-point scale (e.g., "I strive for excellence in everything I do."). The statements were taken from the Conscientiousness section of the NEO-FFI. The dependent measure is the sum of answers (with some items reversed in scoring), with a lower score indicating a self-perception as less conscientious.

Short Grit Scale (Angela Lee Duckworth & Quinn, 2009) – Participants were asked to rate how well-described they were by 8 statements assessing their perception of their own "grit" on a 5-point scale (e.g., "I am diligent."). The dependent measure is the sum of their answers (with some items reversed in scoring), with a lower score indicating a lower self-perception of grit.

Training protocols

For both the dual n-back and MOT groups, training sessions lasted approximately forty minutes per day, and participants were asked to commit to one training session per day, Monday through Friday, at a consistent time. In the event that a training session was missed, participants were allowed to train on the weekend, or to train twice in one day, so long as the two sessions were separated by at least three hours of time. This option was used by 3 of the MOT participants (with a maximum of 3 double-session days) and 6 of the n-back participants (one subject had double-sessions on 5 days in an attempt to complete the experiment before winter break, the other five had a maximum of two double-sessions). Participants in the dual n-back training group completed 20 sessions in an average of 29.2 days (min 21 days, max 42 days), while participants in the MOT training group completed 20 sessions in an average of 28.6 days (min 23 days, max 37 days).

In addition to the weekly bonus payment for completing all five sessions in that week, participants were emailed on a weekly basis congratulating them on their attendance, alerting them of their bonus, and informing them of the progress they had made in training that week. This email was intended to be motivational, so the email highlighted new achievements from the previous week (e.g., a new peak in a performance measure).

Multiple Object Tracking (MOT) Training Participants assigned to the MOT task performed 90 adaptive tracking trials per day, as described in the baseline MOT testing session. (Due to experimenter error, three MOT participants had some days of training with 60 trials instead of 90 trials. These days were during the first half of the training period, and no subject had more than three short days.) The initial speed of the tracked objects was determined by the final speed of the pre-training baseline MOT session, which was reached via the staircasing procedure described above. On subsequent days, the first trial's speed was set to the speed of the last trial on the previous training day. The speed of the tracked objects was adjusted upward by .5 degrees of visual angle/second whenever two consecutive trials were answered correctly and downward by .5 degrees/second when two consecutive trials were missed. Participants were allowed to take breaks, as needed.

Dual n-back Training Participants assigned to the dual n-back training group performed 30 blocks of the task per session, as described above. Due to evidence for a dose-dependent relation between the amount of dual n-back training and gains in transfer to fluid intelligence in Jaeggi et al., 2008 (Jaeggi et al., 2008), we provided all participants with more training (30 blocks/session) than the highest level of training in that study (20 blocks/session) to maximize the dose of training received and to increase the likelihood that the WM training would yield near- and far-transfer gains.

The manner in which the difficulty adapted followed the task described in Jaeggi, et al, 2008: If the participant made more than 5 errors in a block, the *n* of the next block was decreased by 1, to a minimum of a 1-back block. If the participant made 2 or fewer errors in both the auditory and visual n-back stream, the *n* of the next block was increased by 1, with no maximum *n-back* level. In all other cases, the difficulty level remained the same. All participants started at a 2-back level on day 1, and on later days their starting difficulty was set to be the same as the last block on the previous day. Participants were encouraged to take short breaks, as needed, to stay focused during training.

Analyses of the pre- and post-training n-back measures and of the training n-back sessions had to be conducted with different dependent measures. The pre- and post-training measures were analyzed with accuracy (d') because loads were held constant. The training sessions could be not analyzed in this way because load was adaptively altered to keep accuracy as constant as possible. Therefore, training session measures were analyzed with load as the dependent measure.

Results

Initial Group Comparisons

One-way ANOVAs confirmed that the groups did not significantly differ on any of the behavioral measurements (all p's > .19 for 3-group comparison, all p's > .19 for comparison between the two active training groups).

Initial Task Correlations

One plausible reason to expect transfer from a trained task to an untrained task is that those tasks share common cognitive or neural processes, as evidenced by high correlations between those tasks. The full set of correlations between initial scores on the behavioral outcome measures and initial scores on the training tasks is displayed in Table 2. Of particular importance for the hypothesized transfer from the dual n-back task to more general cognition, a participants' initial performance on the dual n-back task (as measured by d' across the 2-back to 6-back difficulty levels) was significantly correlated with fluid intelligence measures (Ravens Advanced Progressive Matrices and the WASI Matrix Task), complex working memory measures (Operation Span, Reading Span, and their combined score), and a reading comprehension measure (Nelson-Denny Reading Comprehension). In contrast, the adaptive control task (MOT) did not show significant correlations with those measures.

Table 2. Initial Task Correlations with Training Tasks

Behavioral Task	Correlation with Initial Dual n-Back d'	p-value	Correlation with Initial MOT Speed	p-value
Initial MOT Speed[a]	0.19	.149	N/A	N/A
<u>*Complex Working Memory Measures*</u>				
Operation Span Score[a]	0.36	0.006	0.26	0.055
Reading Span Score	0.27	0.043	0.14	0.312
Combined Span Score[a]	0.36	0.006	0.22	0.100
<u>*Fluid Intelligence Measures*</u>				
RAPM Score (out of 17)	0.50	<.001	0.19	0.160
<u>*Weschler Abbreviated Scale of Intelligence Subtests*</u>				
WASI Blocks	0.42	<.001	0.41	0.001
WASI Matrices	0.28	0.033	0.10	0.479
WASI Similarities	0.23	0.086	0.08	0.540
WASI Vocabulary	0.24	0.073	0.16	0.222
<u>*Reading Measures*</u>				
Nelson Denny Reading Rate	0.16	0.241	0.13	0.337
Nelson Denny Comprehension	0.28	0.031	0.24	0.069
<u>*Speed of Processing Tasks*</u>				
Woodcock Johnson III Pair Cancellation	0.29	0.029	0.28	0.033
Woodcock Johnson III Visual Matching	0.18	0.170	0.46	<.001
Digit/Symbol Coding	0.21	0.112	0.23	0.081
<u>*Personality Measurements*</u>				
Conscientiousness[a]	-0.03	0.818	0.01	0.938
Dweck	-0.10	0.438	-0.03	0.829
Grit	0.117	0.384	0.04	0.795

Table 2. Pre-training correlations with to-be-trained tasks.

Correlations between initial scores on the two training tasks and the behavioral outcome measures are shown. Statistically significant (uncorrected for multiple comparisons) correlations are bolded. Unless otherwise specified, correlations are across 58 participants (19 passive control, 19 multiple object tracking, 20 dual n-back). [a] – 19 dual n-back measurements.

Trained Tasks

Both active training groups improved significantly with practice on their trained task (Figure 1). In the dual n-back training condition, participants improved from an average n-back of 3.19 (SD = .47) over the first three days of training to an average n-back of 5.1 (SD = 1.1) across the last three days of training [$t(19) = 9.70$, $p < .0001$]. All 20 dual n-back participants improved substantially, with everyone completing at least one dual 5-back block, 17 participants completing a dual 6-back block, 12 participants completing a dual 7-back, 6 participants completing a dual 8-back, and 3 participants completing one or more dual 9-back blocks. (Figure S1 shows individual training gains for both the MOT and dual n-back groups.)

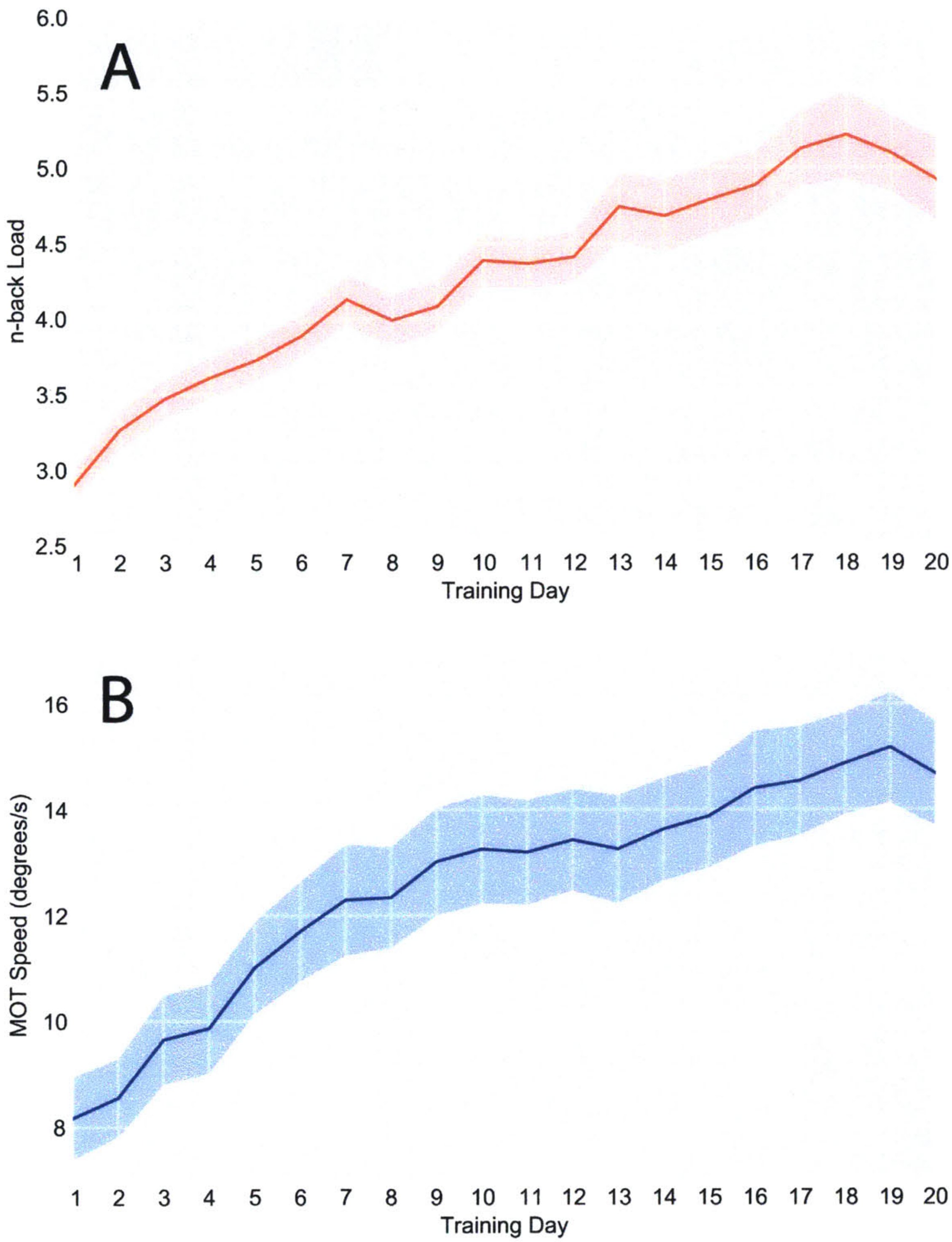


Figure 1. Performance across training sessions. A) Mean dual n-back load and B) mean multiple object tracking speeds achieved per session of training are displayed. Shaded area represents standard error of the mean.

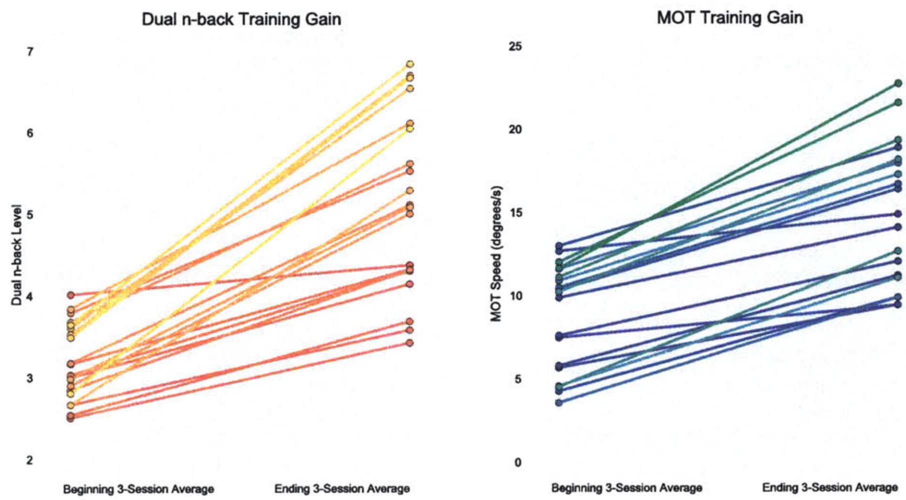


Figure S1. Individual Subject Training Gains. Beginning and ending dual n-back loads/Multiple Object Tracking (MOT) speeds are presented for each participant. Beginning points represent the average performance across the first three days of training, while ending points display the average performance across the final three days of training.

Some participants reported changing their strategies for performing the dual n-back task throughout training. The most commonly reported strategy was mentally superimposing the auditorily presented letter in the visually presented spatial location in an attempt to consolidate the two input streams, though this was not a universally reported strategy. However, the improvement observed on this task cannot be explained merely by strategy shifts. Participants reported fixing on their own idiosyncratic strategies during the first few days of training, and continued making improvements over the course of the 20-day training period long after their particular strategy was chosen.

In the MOT training condition, participants improved from an average tracking speed of 8.8 degrees/second (SD = 3.2) over the first three days of training to an average speed of 14.9 degrees/second (SD = 4.2) over the last three days. [$t(18) = 11.6$, $p < .0001$] Although there was a range of improvement in this condition, all participants were able to track items at least 12 degrees/second at some point during their training, with six participants becoming able to track 4 targets moving at faster than 20 degrees/second.

Some MOT participants also reported changing strategies early in the course of training, although these participants tended to fixate on a strategy early in training and then continued using it throughout the remainder of the training period. Strategies varied widely, with the most commonly reported three strategies being (1) to visualize the tracked dots as corners of a quadrilateral, (2) to attempt to track the center of mass of the four target dots, or (3) to remain fixated on the center fixation cross and track all four target dots in the periphery, without trying to merge the targets into a coherent single object.

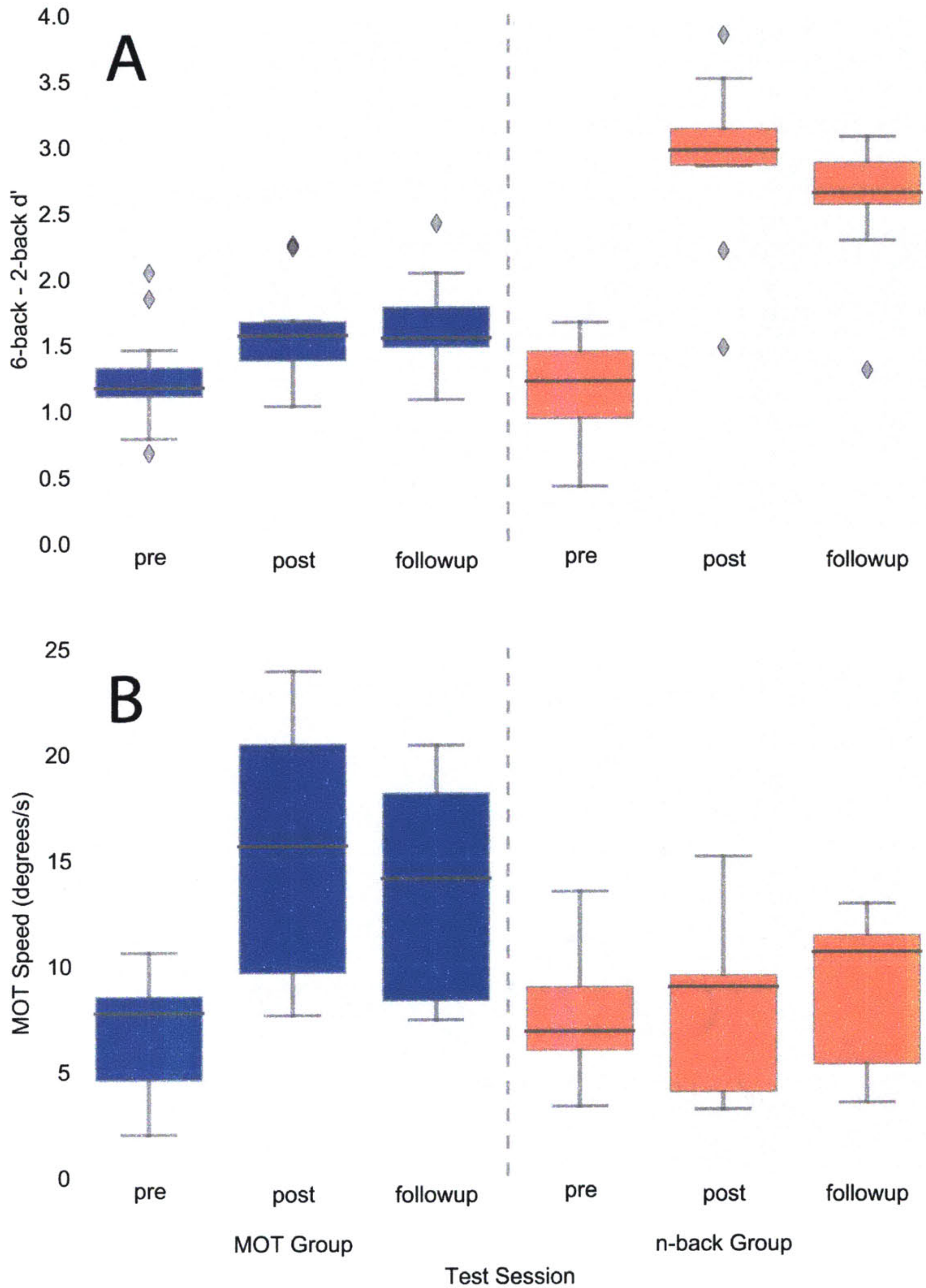
Improvements on both the n-back and MOT tasks were specific to their training group. Comparing performance on these two tasks during the behavioral testing before and after training reveals a double-dissociation between the groups – the MOT training group improved on the pre- and post-training MOT task significantly more than did either the passive control or the n-back group [Group x Time interaction, $F(2,117) = 37.7$, $p < .0001$] , while the n-back group improved on the n-back task significantly more than either the passive control or the MOT training groups [Group x Time interaction,

$F(2,117) = 47.3$, $p < .0001$]. Direct comparison of the two training groups with the No-Contact group revealed whether either training group exhibited any transfer to the untrained task. The MOT group exhibited no more gain on the dual n-back task than the No-Contact group [Group x Time interaction, $t(55) = .17$, $p = .86$], and the dual n-back group exhibited no more gain on the MOT task than the No-Contact group [Group x Time interaction, $t(54) = .56$, $p = .57$).

Duration of Training Gain

Gains made on a trained task largely persisted for 6 or more months past the end of training (Figure 2). Comparing the d' measurement for the 2-back through 6-back tests for the group of participants who received n-back training, paired t-tests showed significant improvements from pre-training testing (M: 1.23, SD: . 42) to post-training testing (M: 2.92, SD: .67), [$t(9) = 7.49$,$p < .0001$], and from pre-training testing to follow-up testing (M: 2.60, SD: .53) [$t(9) = 7.96$, $p < .0001$]; there was also a significant decrease from post-training testing to follow-up testing [$t(9) = 3.57$, $p < .01$]. In comparison, the MOT training group showed a smaller gain from pre-training testing (M: 1.25, SD: .40) to post-training testing (M: 1.59, SD: .39) [$t(10) = 4.96$, $p < .001$], but no significant difference between the post-training testing and the follow-up testing (M: 1.64, SD: .36) [$t(10) = .96$, $p > .35$].

For the MOT speed assessment, a similar pattern of enduring skill emerged (Figure 2). The MOT training group showed significant improvements from pre-training testing (M: 6.60, SD: 2.87) to post-training testing (M: 15.43, SD: 5.92) [$t(10) = 8.03$, $p < .0001$], and from pre-training testing to follow-up testing (M: 13.75, SD: 5.08) [$t(10) = 9.27$, $p < .0001$]; there also was a significant decrease from post-training testing to follow-up testing [$t(10)= 3.13$, $p< .05$]. In comparison, the n-back training group did not show a significant gain from pre-training testing (M: 7.20, SD: 3.1) to post-training testing (M: 8.37, SD: 3.82) [$t(8) = 1.43$, $p > .18$], or from pre-training testing to follow-up testing (M: 8.87, SD: 3.45) [$t(8) = 1.46$, $p > .18$]. The post-training testing was also not significantly different than the follow-up testing for the n-back training group [$t(9) = 1.02$, $p > .33$].



A
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
6-back - 2-back d'
pre
post
followup
pre
post
followup
B
25
20
15
10
5
0
MOT Speed (degrees/s)
pre
post
followup
pre
post
followup
MOT Group
n-back Group
Test Session

Figure 2. Duration of training effects. A) Difference between dual 6-back d' and dual 2-back d' is shown for pre-training, post-training, and six-month follow-up sessions for both active training groups. **B)** Multiple object tracking speed is shown at all three time points for both active training groups. Solid dark, horizontal line indicates condition median; filled areas encode middle 50%. Whiskers extend 1.5 times the interquartile range beyond the box bounds.

Transfer Tasks

In contrast to the substantial improvements seen on the trained tasks, participants did not generally show improvements on the tasks measuring near or far transfer (Table 3). The one statistically significant improvement (although it did not survive Bonferroni correction for multiple comparisons) was on the Matrix Reasoning section of the Wechsler tests, where the MOT group showed an average improvement of 2 items that was not observed in the n-back training group.

Task	n-back pre-test score (SEM)	n-back post-test score (SEM)	MOT pre-test score (SEM)	MOT post-test score (SEM)	Control pre-test score (SEM)	Control post-test score (SEM)	n-back/MOT training interaction p-value	3-way interaction p-value	Minimum Detectable Effect Size (sensitivity)
Trained Tasks									
d' (2-6 back)	**1.09 (.09)**	**2.60 (.16)**	**1.14 (.12)**	**1.38 (.15)**	**1.15 (.11)**	**1.41 (.14)**	**< .0001**	**< .0001**	**.19**
MOT Speed[a]	**6.28 (.63)**	**7.19 (.85)**	**6.76 (.68)**	**14.3 (1.31)**	**6.13 (1.0)**	**6.36 (.96)**	**< .0001**	**< .0001**	**.08**
Complex Working Memory Measures (Near Transfer)									
Operation Span[b]	47.4 (3.15)	58.5 (2.52)	47 (3.06)	51.7 (2.7)	51.6 (4.0)	60.1 (3.28)	.210	.384	.19
Reading Span[c]	48 (3.06)	52.4 (2.68)	41.8 (3.4)	39.3 (3.38)	46.1 (4.10)	47.4 (4.48)	.176	.306	.11
Combined Span[b]	95.9 (5.69)	110.8 (4.66)	88.8 (6.06)	91 (5.36)	97.8 (6.64)	107.5 (7.15)	.154	.267	.13
Fluid Intelligence Tasks (Far Transfer)									
RAPM	13.3 (0.47)	13.2 (0.67)	13.6 (0.46)	13.3 (0.5)	13.3 (0.49)	12.7 (0.62)	.827	.861	.29
WASI/WAIS Subtasks (Far Transfer)									
Vocabulary	14.5 (0.48)	16.1 (0.6)	14.3 (0.33)	15.9 (0.45)	13.7 (0.44)	15.4 (0.34)	.871	.983	.27
Blocks	13 (0.35)	13.9 (0.57)	13.3 (0.45)	14.6 (0.59)	12.2 (0.52)	13.3 (0.61)	.464	.737	.16
Similarities	13 (0.45)	14.2 (0.49)	13 (0.4)	13.8 (0.51)	13.3 (0.25)	14.1 (0.46)	.558	.748	.29
Matrix Reasoning	**13 (0.4)**	**13.4 (0.6)**	**12.9 (0.37)**	**14.7 (0.49)**	**12.6 (0.45)**	**13.7 (0.45)**	**.024**	**.096**	**.32**
Nelson-Denny Reading Measurements (Far Transfer)									
Comprehension[d]	236 (2.81)	236 (2.58)	237 (2.25)	237 (2.62)	240 (1.88)	240 (1.97)	.980	.979	.32
Reading Rate[d]	209 (4.1)	216 (4.53)	213 (5.82)	221 (5.74)	214 (5.08)	216 (4.49)	.684	.447	.19
Speed of Processing Tasks (Far Transfer)									
Digit Symbol Coding	11.8 (0.59)	13.3 (0.64)	10.9 (0.64)	11.9 (0.65)	11.4 (0.51)	12.6 (0.48)	.421	.694	.20
Visual Matching[d]	105 (2.2)	109 (1.95)	106 (3.75)	110 (2.81)	105 (3.17)	106 (3.44)	.851	.273	.14
Pair Cancellation[d]	98.8 (2.34)	105 (2.47)	97.6 (2.77)	102 (2.66)	97.7 (2.44)	104 (2.29)	.671	.797	.29

Table 3. Transfer from Trained Tasks. Pre- and post-testing means and standard errors of the means are presented for each treatment group. The interaction terms from repeated-measures ANOVAs show significant differences between treatments. Statistically significant (uncorrected for multiple comparisons) results are bolded. Unless otherwise specified, analyses include 19 passive control participants, 19 MOT participants, and 20 dual n-back participants. [a] -19 dual n-back. [b] -14 passive control, 19 dual n-back. [c] – 14 passive control. [d] – 18 passive control.

Power Analyses

To assess whether the observed lack of transfer was a result of an underpowered sample size, we used the G*Power software package (Faul, Erdfelder, Lang, & Buchner, 2007) to assess the sensitivity of the behavioral tests. The final column of Table 3 reports the minimum effect size that could be detected in a between-groups interaction, based on the sample sizes in this study and the correlation between the pre- and post-testing scores for each test. (This test uses the correlation between the pre- and post-training scores in the passive control group as a measure of test-retest reliability. The more consistent the relationship between the two scores is, the smaller the detectable change will be.) The sensitivity level was set at $p < .05$ for failing to observe a real effect. For every transfer measure, this experiment had sufficient power to detect a medium ($f = .25$) to large ($f = .40$) effect, and had ample power to detect the effect sizes reported in the initial Jaeggi experiment ($d = .68$).

Correlations between Training Improvement and Transfer

Some prior research has observed transfer gains in only those participants who successfully improved on the trained task (Jaeggi et al., 2011; Novick, Hussey, Teubner-Rhodes, Harbison, & Bunting, 2013). We therefore performed several analyses to examine whether there were individual differences among participants that were associated with either training gains or with transfer from training to other measures. For this purpose, a training improvement score was calculated for each participant by subtracting the average performance during the initial three days of training from the average performance during the last three days of training. For the n-back training group, the average "n" of the n-back blocks was calculated, whereas for the MOT training group, the average object movement speed was calculated.

One method of assessing whether the amount of training improvement affects the degree of transfer is to measure the correlation between training and transfer gains. For both the n-back and MOT groups, a positive correlation was observed between the amount of improvement during training and the amount of improvement on the trained task between the pre- and post-assessment (n-back $r = .85$, $p < .0001$; MOT $r = .77$, $p < .0001$). However, the amount of training gain did not significantly predict improvement on any transfer task; participants who improved to a greater extent on the training tasks did not improve more or less on potential transfer tasks than did participants who improved to a lesser extent (all n-back r values < .33, all p's > .15; all MOT r values < .38, all p's > .11). Figure S2 depicts the absence of a relation between improvement on trained tasks and the post-training changes in the RAPM and the combined span tasks.

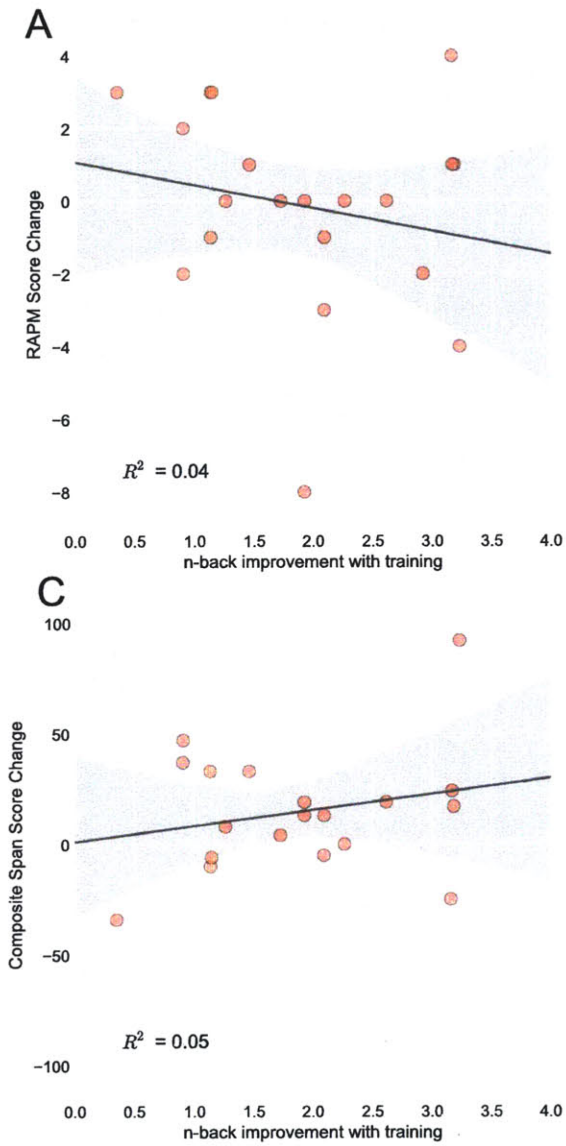


Figure S2. Relationships Between Training Gains and Transfer Measures. A) Correlation between improvement on the dual n-back task during training and the difference between pre- and post-training Ravens Advanced Progressive Matrices (RAPM) scores. **B)** Correlation between dual n-back improvement and change on the Composite Span Task scores. **C)** Correlation between improvement in Multiple Object Tracking (MOT) speed and RAPM change. **D)** Correlation between MOT gains and Composite Span Task score changes. All p's > .05. Error bands are bootstrapped 95% confidence intervals for the regression.

Another analysis that has previously revealed a difference in transfer between participants who exhibited larger or smaller training gains has been a division of participants into groups based on training gains above or below the group median (median split) (Jaeggi et al., 2011). Such a median split of participants in the present study who performed the n-back training yielded no significant differences in transfer between groups (all n-back t-ratios < 1.78, all p's > .09). The only transfer measure that approached significance (at p=.09) was on the RAPM test, in which the participants who improved *less* on the trained n-back task had higher scores on the post-training behavioral testing. Similarly, when separating the MOT participants into two groups based on median MOT improvement, the two groups showed no significant differences in transfer performance (all MOT t-ratios < 1.74, all p's > .10).

A clustering algorithm (an example of which is the k-means algorithm (Jones, Oliphant, & Peterson, 2001)) is another approach to classifying participants into two groups based on differences in training gains, and this approach has shown that participants classified as responding to training show gains on transfer tasks, whereas participants classified as not responding to training fail to show gains on transfer tasks (Novick et al., 2013). Clustering algorithms have the advantage of classifying different numbers of participants into responder and non-responder groups when such a division does not occur naturally at the median. The clustering algorithm applied to the present n-back training data yielded two clusters, one with 9 participants and the other with 11 participants, which was a close approximation of the median split grouping that had yielded 10 participants in each group, and again there were no significant differences revealed in the transfer measures between groups (all t-ratios < 1.50, all p's $> .15$).

Clustering algorithms, however, do not always yield meaningful or easily interpretable clusters. The same clustering algorithm applied to the MOT training data yielded clusters with 14 "non-responder" participants and 5 "responder" participants, although the average participant in the "non-responder" cluster increased their MOT tracking speed by more than 5 degrees/second.

Correlations between Pre-training Measurements and Training Gains

No pre-training behavioral score significantly predicted the amount of task improvement during training in either the n-back or the MOT group (all n-back r values $< .31$, all p's $> .17$; all MOT r values $< .34$, all p's $> .16$).

Correlations of Personality Measurements and Transfer

We also examined whether personality assessments were associated with different training or transfer outcomes. Neither the Dweck measure of attitude toward intelligence (a "growth mindset") nor measures of conscientiousness or grit correlated significantly with training gains on either training task, although there was a trend toward a significant negative correlation between the growth mindset and improvement on the n-back training task ($r = -.44$, $p = .051$), such that participants who viewed intelligence as more malleable had less improvement across their n-back training. A greater growth mindset score was positively correlated, however, with improvement on the Ravens Advanced Progressive Matrices in the n-back group ($r = .53$, $p = .017$) and in the passive control group ($r = .51$, $p = .027$), but not in the MOT control group ($r = .031$, $p > .9$). No other transfer measures were significantly predicted by growth mindset scores.

Although the conscientiousness scores and "grit" scores were highly correlated in each of the three treatment groups (n-back $r = .75$, $p < .001$; MOT $r = .70$, $p < .001$; passive $r = .76$, $p < .001$), the two measures differed in their correlations with the behavioral outcome measures. A higher "grit" score predicted less improvement on the RAPM for the n-back group ($r = -.45$, $p = .049$) and the MOT group ($r = -.58$, $p = .009$), such that participants who viewed themselves as having more "grit" improved less on the RAPM after training, although this relationship did not hold for the No-Contact group ($r = .17$, $p = .5$). Similarly, a higher score on the conscientiousness measure predicted less improvement on the RAPM for the MOT group ($r = -.57$, $p = .01$), such that participants who saw themselves as more conscientious improved less on the RAPM after training, although this was not observed in either of the other two groups (n-back $r = -.21$, $p = .37$; no-contact $r = -.04$, $p = .85$). Finally, a high conscientiousness score predicted a lower Pair Cancellation improvement within the MOT group ($r = -.47$, $p = .04$), but not in the n-back or no-contact control groups (n-back $r = -.07$, $p = .77$; no-contact $r = -.13$, $p = .58$). No other transfer measures were significantly predicted by either conscientiousness or grit scores.

Discussion

This experiment yielded one major finding and some new observations. The major finding was a failure to observe any gains in measured fluid intelligence after working memory training. Although participants improved substantially on their trained tasks, neither WM training nor multiple object tracking training provided benefits on speed of processing tasks, other standardized measures of intelligence, or measurements of reading comprehension. The lack of transfer from WM training to other measures occurred for both near-transfer tasks (other complex working memory tests) and far-transfer tasks (e.g., fluid intelligence measures) and was relative both to an active control training group (MOT training) and a no-contact control group. The absence of transfer occurred despite robust learning on the trained tasks and substantial retention of those acquired skills lasting over six months.

Magnitude of Training Effects

Critically, the amount of improvement seen on the dual n-back task was nearly identical to the amount of training improvement seen in the prior study reporting improvements in fluid intelligence (Jaeggi et al., 2008). In the previous report, participants initially were able to perform a dual 3-back task, and ultimately averaged slightly better than a dual 5-back task after 19 days of training. In the present experiment, participants' average performance across the first three days was 3.19-back, and average performance across the last three days was 5.19 back. Participants in previous attempts to replicate the original Jaeggi finding exhibited lesser amounts of dual n-back improvement across training. Specifically, participants achieved an average dual n-back level of approximately 4.0 in one study (Redick et al., 2012), and an average of approximately 4.1 in another study (Chooi & Thompson, 2012). The somewhat lower final levels of WM performance in these two failed replications left open the possibility that the discrepent findings on transfer to fluid intelligence were related to the level of WM capacity learned through training. The present WM training outcomes, which closely resemble those from Jaeggi et al., 2008, indicate that failure of transfer to other measures of cognition and fluid intelligence cannot be accounted for either by gains in trained WM performance or in final level of WM performance.

Although participants in the present study had 33% more training per session than the those in the study from Jaeggi et al, 2008, they did not exhibit greater gains in WM capacity than those who had an equal number of sessions in the Jaeggi et al., 2008, study. It is not clear why the additional training did not yield additional WM capacity. One possibility is that there is a limit or asympote to dual n-back training. Another possibility is that participants vary in the rate of WM training gains, which could be related also to transfer.

The amount of improvement on the active-control MOT task was comparable with the amount of improvement observed in the dual n-back group. Participants, on average, improved their initial score by 1.59x in the dual n-back condition, while participants trained in the MOT condition improved their initial score by 1.69x, from an initial 3-day average of 8.8 degrees per second to a final 3-day average of 14.9 degrees per second. This comparable level of improvement validates the use of the MOT task as a suitable active control for the dual n-back task. Although the MOT task has been widely used to study visuospatial WM capacity (e.g., (Alvarez & Franconeri, 2007)), this is the first study to show that MOT skill can be acquired and maintained over a long period.

Specificity and Duration of Training Effects

Training in both active groups was robust and specific to the type of training in that participants who were trained on the one task exhibited substantial gains on the trained tasks, but no gains on the other task. The duration of sustained improvement from dual n-back or MOT training, however, has been previously unknown in healthy young adults. In this experiment, 18 participants returned after their 20 days of training to assess the longevity of their specific training gains. Both the MOT and n-back groups showed significant improvement from their pre-testing to post-testing scores, and those improvements were largely, although not completely, maintained 6 months later. Although we failed to observe improvements in fluid intelligence in this experiment, the maintenance of the training improvements, despite 6 months without further training, seems to be a necessary component of any working memory training paradigm aimed at creating enduring improvements.

Transfer – General Expectations

Transfer from a trained task to an untrained task is expected when the two tasks share common components, whether they be cognitive processing steps or reliance on similar neural activations (Schunk, 2007). In *near-transfer* tasks, the trained task bears surface similarities to the target task, such that observed improvements on the target task could conceptually be the result of either a learned strategy during training that is also applicable to the transfer task, or the result of actually improving an underlying cognitive skill. In *far-transfer* tasks, the demands of the task do not involve an overt shared strategy, so there are fewer mechanisms for training in one task to produce benefits on the second. Although experiments examining transfer from the dual n-back task to fluid intelligence (*far transfer*) have reported mixed results (Chooi & Thompson, 2012; Jaeggi et al., 2010, 2008, 2011; Redick et al., 2012; Studer-Luethi et al., 2012), some WM training studies report transfer from the trained WM task to another untrained WM task (*near transfer*) (e.g., (Brehmer et al., 2012; Buschkuehl et al., 2008)), including one study in which dual n-back training similar to that used here resulted in improved operation span performance (Anguera et al., 2012).

Near Transfer

In the present experiment, the two tasks most conceptually similar to the dual n-back training were the Operation Span and Reading Span tasks that, like the dual n-back task, are tasks of complex working memory (CWM). The finding that performance on all three CWM tasks was significantly correlated supports the idea that the three tasks share underlying mechanisms. Previous experiments training WM tasks have sometimes shown transfer to non-trained WM tasks (e.g., (Anguera et al., 2012; Brehmer et al., 2012; Buschkuehl et al., 2008)). In this study, however, there was no evidence of transfer from the dual n-back task to the Operation or Reading Span tasks alone or in combination, or in relation to the amount of learning on the n-back task.

Far Transfer - Fluid Intelligence

The possibility that WM training would enhance fluid intelligence is supported by behavioral findings reporting high correlations between complex WM scores and fluid intelligence scores (Andrew R.A. Conway et al., 2003; Randall W. Engle, Tuholski, Laughlin, & Conway, 1999), which indicates shared psychological mechanisms, and neuroimaging findings reporting similar activations for complex WM and fluid intelligence tasks, which indicates shared neural mechanisms (Burgess, Gray, Conway, & Braver, 2011; Andrew R.A. Conway et al., 2003; Gray, Chabris, & Braver, 2003; Kane & Engle, 2002). Indeed, we also observed strong correlations between initial performance on the dual n-back task and two measures of fluid intelligence. This relationship was specific -- there was no correlation between initial performance on the MOT task and the same measures of fluid intelligence.

There was not, however, any improvement on the fluid intelligence tasks after dual n-back training compared to either the active control group or the passive control group. There was also no relation between the amount of improvement on the dual n-back task and transfer to either fluid intelligence measure. Thus, although it appears that the necessary conditions for transfer to occur were achieved in the experiment, there was no evidence of transfer from WM training to fluid intelligence measures.

There was one significant transfer effect from training to a fluid intelligence measure: MOT training improved performance on the matrix reasoning section of the Weschler Intelligence Tests. Although this finding may be of interest, there are two reasons to suspect it could be spurious. First, the group who received MOT training showed no more improvement in the matrix reasoning score than did the no-contact control group. Second, the improvement by the MOT-trained group did not extend to the other matrix-based fluid intelligence measure, the Ravens' Advanced Progressive Matrices. For these reasons, as well as the absence of any behavioral correlation between initial MOT performance and either measure of fluid reasoning, it seems more likely that this was an example of the sort of false positive finding that can occur with so many behavioral measures, rather than genuine transfer from MOT training to fluid intelligence.

Far Transfer – Other Tasks

WM training has sometimes been reported to yield transfer to other kinds of performance, including improvements in domains of cognitive control that are often associated directly with WM, such as attentional control (e.g., Stroop task) and reading comprehension (Chein & Morrison, 2010), among others (reviewed in (Melby-Lervåg & Hulme, 2013)). These findings motivated inclusion of additional measures, including the specific test of reading comprehension that demonstrated benefit from WM training (Chein & Morrison, 2010) and processing speed (which typically correlates with WM capacity) (Andrew R.A Conway, Cowan, Bunting, Therriault, & Minkoff, 2002). We did not, however, observe transfer from either WM or MOT training on any of these measures.

Working Memory in the Multiple Object Tracking Task

The MOT task was conceptualized as a control training task involving perceptual skill learning, but learning on the MOT task could alternatively be conceptualized as training of visuospatial working memory. The fact that pre-training MOT performance did not correlate with complex working memory tasks, matrix reasoning tasks, or reading comprehension measures indicates that if MOT involves working memory, it may selectively involve the visuospatial component and not executive or phonological components. Further, the substantial gains on MOT performance did not produce transfer to other tasks, with the exception of a single isolated measure.

Personality Measurements and Motivation

There is evidence that personality factors can modulate the influence of WM training on gains in fluid intelligence (Studer-Luethi et al., 2012), and we examined personality factors that could, in theory, influence such transfer. One study found that greater conscientiousness predicted higher levels of performance during training on single n-back tasks, although it did not predict performance in a separate group using dual n-back training (Studer-Luethi et al., 2012). Furthermore, across both n-back training groups, conscientiousness was negatively correlated with fluid intelligence gains. In contrast, we did not observe a correlation between conscientiousness (or the highly correlated "grit" scores) and performance during either dual n-back training or MOT training. Similar to the prior findings, higher conscientiousness scores predicted smaller improvements on the Ravens Advanced Progressive Matrices (RAPM) in the MOT training group, while higher grit predicted smaller improvements on the RAPM in both the dual n-back and MOT training groups. We fail to support their broader claim that conscientiousness negatively predicts transfer to fluid intelligence, however, as neither conscientiousness nor grit scores predicted change in the performance IQ measures of the WASI/WAIS test (the matrix reasoning task and block design measures) that load highly on fluid intelligence.

Another plausible variable affecting transfer from WM training to fluid intelligence is the participant's attitude about intelligence. In some studies, students who believe that intelligence is a malleable trait that can be enhanced by effort (i.e., student who have a "growth mindset") show greater learning than students who believe that intelligence is a fixed trait (Blackwell et al., 2007). This study did not show that pattern. Instead, we observed a trend in the opposite direction for the n-back training group – participants who viewed intelligence as fixed improved more over the course of n-back training than did participants with a growth mindset toward intelligence. We observed no relation between attitudes toward intelligence and improvement on multiple object tracking.

There were also some relations between growth mindset and improvement on the RAPM. Participants with greater growth mindsets in the n-back group exhibited greater growth on RAPM scores. Although this could be interpreted as revealing that greater growth mindset facilitates greater transfer of working memory training to fluid intelligence, two other findings contradict this interpretation. First, greater growth mindset was not related to gains on the other fluid intelligence measure or on other working memory tasks. Second, the same relation between greater growth mindset and greater growth on RAPM scores was observed in the No-Contact group who had received no training. Personality did not seem to account for variation in transfer from WM to other kinds of cognitive ability.

A more general measurement of motivation is difficult to obtain. It is possible that we did not observe the same benefit of WM training for fluid intelligence as previous groups because our participants were somehow less motivated to improve. Subjectively, participants appeared excited about the prospect of transfer from WM training to other aspects of their life, especially academic endeavors. Certainly, every effort was made to motivate the trained participants in both the dual n-back training and the active control groups through several means: 1) by explicitly telling both groups that the task could possibly make them "smarter", 2) by providing weekly encouraging e-mails highlighting their accomplishments, and 3) by providing monetary bonuses for conscientious training. Although it is unknown how effective these manipulations were, the overall amount of training improvement seen in this study was nearly identical to that seen in the original Jaeggi study, providing at least an indirect confirmation of similarly motivated participants.

Sensitivity to Detect Transfer

Failure to observe transfer could reflect insufficient statistical power, but for several reasons it appears that this unlikely to explain the lack of transfer observed in the present study. First, there were almost no transfer effects in the training groups that numerically surpassed the simple test/retest practice effects exhibited by the no-contact control group. Second, the effect size in the initial report of transfer from WM training to fluid intelligence was substantial (Cohen's d = .65) (Jaeggi et al., 2008). The sample in the present study would have allowed detection of transfer with an effect size of d =.27 or better. In the social sciences, a "small" effect size for an independent means t-test (which is the statistic that the interaction of a repeated measures ANOVA evaluates) is regarded as d = .2, while a "medium" effect size on this test is d =.5 (J. Cohen, 1992). Therefore, the present study ought to have had sufficient power to replicate the initial report and to find most small effects, although it may have been underpowered to detect very small differences between training groups.

Implications for Working Memory Training

The goal of enhancing core cognitive abilities that support and constrain performance in many cognitive domains is an important educational and clinical goal, and speaks to basic theoretical interests about plasticity of the human mind and brain. For these reasons, the report that WM training enhances fluid intelligence (Jaeggi et al., 2008) has generated great interest for many researchers in human psychology and human cognitive neuroscience as well as in the public at large. The promise of such training for enhancing the cognitive capacity of the human mind has been supported by other studies reporting WM training benefits on reading comprehension (Chein & Morrison, 2010; Loosli, Buschkuehl, Perrig, & Jaeggi, 2012), mathematical ability (Holmes et al., 2009), and ADHD symptomology (Klingberg et al., 2005), and some training programs have even gone so far as to show the neural changes occurring with WM training that theoretically enable the transfer to other domains (McNab et al., 2009; Takeuchi et al., 2010, 2011).

However, several other studies have failed to observe any transfer from WM training to broader cognitive functions. In one well-publicized finding, for example, 11,430 people in the UK performed a variety of on-line cognitive training tasks at home for a 6-week period, and although improvements were found on all trained tasks, there was no near or far transfer to any untrained task (Owen et al., 2010). Reviews and systematic meta-analyses also do not conclude that WM training generally enhances broad cognitive abilities (Melby-Lervåg & Hulme, 2013; Shipstead, Redick, & Engle, 2012).

Broad reviews and discussions about cognitive training often intertwine several distinct issues, such as whether WM training is helpful for young adults, children, older adults with typical age-associated cognitive losses, or patients with diagnoses such as ADHD. The present study, however, focused specifically on the possibility that dual n-back training, if effectively delivered as indexed by gains in WM, could enhance fluid intelligence as reported by Jaeggi et al., 2008. Three published studies, including the present study, have attempted to replicate that finding without success. One study included both an adaptive active control group and multiple measures of both near transfer (WM measures) and far transfer (fluid intelligence, crystallized intelligence, and processing speed measures (Redick et al., 2012)). Like the present study, that research found substantial learning on the trained WM capacity task, but no near transfer to other

WM tasks or far transfer to fluid intelligence, crystallized intelligence, and processing speed. Another study, without an adaptive control group, also failed to find such transfer (Chooi & Thompson, 2012).

It would be valuable to discern the factors across studies that are associated with success or failure in having WM training improve fundamental faculties of the human mind as measured by improved performance on a range of untrained tasks. The present study indicates that the amount of WM training does not appear to account for such variability in outcomes, because the present study involved more training than that reported in a study with positive findings (Jaeggi et al., 2008). Some studies have found that greater gains in training were associated with transfer (e.g., (Jaeggi et al., 2011; Novick et al., 2013)), but we did not observe any such relation between training gains and transfer in three independent analyses. Variation in personality could be another factor (Studer-Luethi et al., 2012), but personality measures of conscientiousness, grit, or attitudes towards intelligence did not correlate with training transfer in the present study. It is difficult at present to identify any one factor across studies that plausibly explains transfer success.

Besides individual differences among participants, another important factor related to transfer gains may be the nature of training program. The present study trained participant on one of two homogenous tasks, the n-back or MOT task. An alternative approach is to employ a training program that involves multiple, heterogeneous cognitive training tasks. Such heterogeneous training has yielded transfer gains in some

WM training studies (e.g., (Jaušovec & Jaušovec, 2012; Novick et al., 2013)). Heterogeneous training may have the advantage of training multiple specific cognitive skills that initially vary across individuals and that promote transfer to heterogeneous transfer tasks that vary in their specific cognitive demands.

It is possible that WM training may be more consistently beneficial for individuals performing suboptimally, rather than the high-performing young adults who have been the participants in the above reviewed studies. There are reports of successful transfer of WM training in patient groups with ADHD (Klingberg et al., 2005, 2002) or stroke (Ostensson et al., 2007), or in particularly younger or older populations (e.g., (Brehmer et al., 2012; Holmes et al., 2009)). It is also these groups of individuals for whom effective cognitive training may be most helpful in improving everyday functioning. Future research will, hopefully, reveal principles by which the effectiveness of cognitive training programs, beyond gains on the trained program itself, can be predicted.

Acknowledgments

The authors thank Aron Milberg for his contributions to statistical analysis, Allyson Mackey for comments on the manuscript, and Rebecca Saxe for her insights and contributions toward the preparation of the manuscript.

Chapter 3: Neural Correlates of an Intensive Working Memory Training Program

Introduction:

Working memory (WM) refers to the maintenance and manipulation of goal-relevant information over a brief time (Baddeley, 1992), and working memory capacity (WMC) refers to the amount of information that can be held in WM. Greater WMC is associated with better performance on a wide range of cognitive tasks, including reasoning, problem solving, and reading comprehension (Daneman & Carpenter, 1980; R W Engle et al., 1999) as well as better academic performance (e.g., Finn et al., 2014; Gathercole, Pickering, Knight, & Stegmann, 2004). Although WMC had been conceptualized as a trait fixed before young adulthood, WM training studies have provided evidence that WMC can be increased in young adults who undergo adaptive WM training (reviewed in Klingberg, 2010). Further, there is some evidence that dual n-back training (which involves the simultaneous encoding, maintenance and updating of both auditory-verbal and visuo-spatial stimuli) can also provide gains on non-trained tasks typically used to measure fluid intelligence, suggesting that improvements from WM training could transfer to meaningful real-world cognitive skills (Jaeggi, Buschkuehl, Jonides, & Perrig, 2008; Schweizer, Grahn, Hampshire, Mobbs, & Dalgleish, 2013 but see also Redick et al., 2012; Thompson et al., 2013) . While the question of far transfer from dual n-back training remains open, participants trained on the task consistently display impressive gains on the dual n-back task itself, typically doubling or tripling their WMC on the task (Thompson et al., 2013). Here, we report the functional brain changes that accompany such a major expansion of task-specific WMC.

Neurophysiological studies of WM in infrahuman primates (Funahashi, Bruce, & Goldman-Rakic, 1989) and neuroimaging studies in humans (e.g., Cohen et al., 1997; Smith & Jonides, 1997) provide convergent evidence that dorsolateral prefrontal cortex (PFC) and parietal cortex support WM operations. PFC and parietal cortices exhibit greater activation under higher loads, when a greater amount of information must be maintained in WM (e.g., Braver et al., 1997; reviewed in Owen, McMillan, Laird, & Bullmore, 2005; Wager & Smith, 2003). Individual differences in WM capacity correlate with differences in activation of PFC and parietal cortices (e.g., Mcnab & Klingberg, 2008). One hypothesis, therefore, is that WM training ought to produce experience-induced plasticity in both PFC and parietal regions.

The PFC and parietal regions activated by tasks that have high WM demands are often described broadly as a “frontoparietal network” (e.g., Owen et al., 2005; Schweizer et al., 2013). While it is clear that that WM tasks provoke activation in the frontal and parietal lobes, it is unclear as to what precise PFC-parietal network supports the trained expansion of WMC. The frontal and parietal lobes include two distinct networks as identified by resting-state analyses of the intrinsic functional organization of the neocortex (e.g., Yeo et al., 2011). One network is the “executive control network” (ECN) that includes the dorsolateral (DLPFC) and dorsomedial (DMPFC) prefrontal cortex and a parietal node centered around the intraparietal sulcus (IPS) (often termed the “frontoparietal network” in the resting state literature). A second, anatomically distinct network is the “dorsal attention network” (DAN) that includes the frontal eye fields

(FEF) and retinotopically-mapped areas in the superior parietal lobe. Here, we examined whether increased WMC involves functional plasticity in one or both of these frontal-parietal networks.

Although functional plasticity associated with WM training has been studied before, there is not a study that has both (1) examined long-term WM training that so powerfully expands task-specific WM capacity, and (2) included an active control condition that allows for distinguishing between functional changes associated with any habituation to a task versus changes associated specifically with enhanced WMC. Prior studies have examined the effects of practice on the neural correlates of WM tasks with single-session practice (e.g., Kelly & Garavan, 2005; Landau, Garavan, Schumacher, & D'Esposito, 2007; Landau, Schumacher, Garavan, Druzgal, & D'Esposito, 2004) or a non-adaptive practice task (e.g., Hempel et al., 2004), but these two situations lack the prolonged and adaptive training that creates the environment necessary to generate expansion of working memory capacity (Klingberg, 2010). Studies that examined adaptive WM training over multiple days (the longest of which was 20 days) are scarcer and have reported disparate findings. Some studies reported increased activation in task-related regions, especially when the pre-training task was initially beyond the ability of participants to perform (Olesen et al., 2004). Others studies reported decreased activation in task-related areas (Schneiders et al., 2012, 2011; Schweizer et al., 2013). Such reduced activation associated with improved performance has been termed as the *neural efficiency* hypothesis (Richard J. Haier et al., 1992), although the neural efficiency explanation has been criticized as a mere redescription, rather than an explanation, of the functional plasticity (Poldrack, 2014).

Here we examined both functional plasticity associated with long-term WM training compared to an active control of long-term visuospatial training matched for intensity and duration. During fMRI, participants performed the dual n-back task at four loads (0-, 1-, 2-, and 3-back) before and after WM training for 40 minutes per day, 5 days per week, for 4 weeks. An active control group who performed a similarly intensive visuospatial skill learning task, multiple object tracking (MOT) (Pylyshyn & Storm, 1988). Participants were randomly assigned to either group, with the constraint that the two groups were equated on initial IQ scores, age, and gender. We examined whole-brain activation changes associated specifically with dual n-back training, examined whether these changes occurred in the executive control and/or dorsal attention networks, and examined whether there were changes in frontoparietal functional connectivity (temporal correlations in activation) that accompanied any activation changes.

Methods:

Participants, Recruitment, and Group Assignment

The 39 participants in this study who underwent an active training task were recruited through web advertisements, physical flyers, and e-mail to the Northeastern and Tufts college mailing lists. Participants were required to be adults between the ages of 18 and 45, right-handed, in good health, and not taking psychoactive medication. All participants provided informed, written consent before participation. This study was approved by the Massachusetts Institute of Technology Institutional Review Board (PI: Leigh Firn).

After recruitment, participants underwent pre-training behavioral testing to determine group assignment. Each participant was paired with another participant based on age, gender, and score on a preselected set of 18 of the 36 problems in the Raven's Advanced Progressive Matrices (RAPM) (Raven, JC et al., 1998) task (as described in Thompson et al., 2013), and each member of that pair was randomly assigned to either the n-back or the MOT training group. An additional 19 participants comprised a passive control group that was not contacted between the first and last behavioral assessments. This group was recruited separately, but in the same fashion, and matched to a training pair by gender and initial RAPM. The three groups did not differ significantly by gender [$X^2(1, N = 37) = .21, p > .65$], RAPM scores [$t(1,37) < 1, p > .48$], or on the full IQ score from the Wechsler Abbreviated Scale of Intelligence (Wechsler, 1999), administered as part of the pre-training battery [$t(1,37) < 1, p > .97$] though the passive control group averaged 1.8 years older than the two active training groups [$F(2,55) = 3.37, p < .05$] (Table 1).

Training Group	Group Size (Num. Females)	Age (STD)	IQ (STD)	RAPM (STD)
MOT	19 (11)	21.3 (2.3)	120.7 (7.0)	13.8 (2.3)
N-Back	20 (13)	21.2 (2.0)	120.9 (10.8)	13.4 (2.1)
Passive Control	19 (12)	23.1 (3.3)	117.6 (7.4)	13.3 (2.2)

Table 1: Participant demographics. IQ measure is the Full 4 IQ measure from the WASI. RAPM measure reflects the number of problems solved on half of the Ravens Advanced Progressive Matrices (see Thompson et al., 2013 for RAPM details).

Fourteen additional potential participants either dropped out of the study or were excluded after initial scanning was completed. Two participants assigned to the dual n-back condition voluntarily withdrew (one after 5 days of training, the other after 9 days); no other participants had begun training when they were excluded or withdrew. The remaining potential participants were not included for a variety of logistical reasons, including difficulties aligning schedules with the experimenters, claustrophobia or excessive movement in fMRI scanning sessions, or repeatedly skipping pre-training appointments.

Participant payment

Participants were paid $20 per training session, with an additional weekly $20 bonus for completing all five training sessions in that week. Participants were paid $20 per hour for the initial and final behavioral testing sessions (approximately 3 hours each) and $30 per hour for neuroimaging sessions (2 hours pre- and post-training of fMRI, 1 hour of electroencephalography). Total compensation for each participant completing the experiment was approximately $900.

Overall experiment design

After recruitment, participants completed a pre-training imaging session which included structural scans for anatomical registration and four runs of the dual n-back task described below. They then completed twenty sessions of adaptive training on either the n-back or MOT task on campus at the Massachusetts Institute of Technology. After training was completed, post-training imaging was administered as quickly as possible. The average number of days between last training session and post-training testing was 4.3 days, with a minimum of 0 days and maximum of 14 days. This time was not significantly different between the two training groups [$t(37) = .2$, $p > .8$]).

Dual n-back Functional Imaging Task Description

Implementation of the adaptive dual n-back training task followed Jaeggi et al., 2008). An auditory letter and a visual square were simultaneously presented for 500ms, followed by a 2500ms response period. Letters were chosen from the consonants B, F, H, J, M, Q, R, and W to maximize auditory discriminability between letters. Squares were presented at one of eight positions evenly spaced around the periphery of the screen. Participants responded when one or both of the current stimuli matched a stimulus presented *n* trials ago. In the "0-back" condition, participants were instructed to respond to a spatial target in the upper-right corner as a visual match, or to the letter "Q" as an auditory match. In order to ensure that each participant fully understood the task, at least one block of each difficulty level was practiced outside the scanner, and participants were allowed to repeat this practice task as necessary until they reported full understanding of the instructions.

Each block presented 10 trials, containing two auditory target trials and two visual target trials, with no trials where both auditory and visual stimuli matched. In order to ensure a consistent level of difficulty within each block, "lure trials" matching either the n+1 or n-1 stimulus were not included. During each block, the current difficulty of the block was indicated at a central fixation point, with additional text labels directly below the difficulty indicator showing the mapping between the two response buttons and "Audio" or "Video" match types.

Responses were made using a scanner-compatible button-box held in the right hand of the participants. A press of the first button, under the participant's index finger, indicated auditory matches and a press of the second button, under the middle finger, indicated spatial matches. No response was required on trials that did not match the target. Participants were allowed the entire 3 seconds following a stimulus presentation to respond to matches.

Each dual n-back run consisted of eight 30-second blocks of the dual n-back task, with 2 blocks of each difficulty from 0-back to 3-back in a counter-balanced order across the four total runs. Each block was preceded by a 3-second instruction screen indicating the n-back load for the upcoming block, and was followed by 16 seconds of rest in order to let the hemodynamic response return to baseline. The total acquisition time for each run was 6 minutes and 36 seconds. Each participant completed 4 runs of the dual n-back task before and after training. In summation, there were 10 trials per block and 8 blocks for each difficulty level, equating to 80 total stimulus presentations at each level. (In 4 cases, scanning delays prevented acquisition of the last run. This occurred once in a pre-training n-back participant, twice in pre-training MOT participants, and once in a post-training MOT participant.)

MRI Data Acquisition

Whole-brain imaging was performed on a 3T Siemens Tim Trio MRI system using a 32-channel coil. Functional images were obtained using a standard T2*-weighted echoplanar pulse sequence (TR = 2 s, TE = 30 ms, flip angle = 90°, 32 slices, 3.0 x 3.0 x 3.1 mm voxels, 20% slice gap, axial slices, interleaved acquisition). Prospective adaptive motion correction was employed to minimize the effects of participant motion. For steady state magnetization, 8 seconds of dummy scans were collected at the beginning of each run, before the experimental paradigm began. Additionally, a whole-brain high-resolution T1-weighted spoiled gradient recalled echo anatomical volume was acquired for purposes of cortical surface modeling, registration to common anatomical space, and across-run alignment. Visual stimuli were projected onto a screen and viewed through a mirror.

Data Analysis

In-scanner Dual n-back Performance

In order to control for response biases between participants, a sensitivity index (Hit Rate – False Alarm Rate) was calculated at each level from 0-back to 3-back for in-scanner performance. Repeated measures ANOVAs were evaluated on this index with training group as a between-subjects factor and task load and session (pre- or post-training) as within-subjects factors.

Functional Imaging Analysis Procedure

Functional imaging data were processed with a workflow of FSL (Smith et al., 2004) and Freesurfer (Dale, Fischl, & Sereno, 1999) tools implemented in Nipype (Gorgolewski et al., 2011) using the lyman interface (http://stanford.edu/~mwaskom/software/lyman/). Each timeseries was first realigned to its middle volume using normalized correlation optimization and cubic spline interpolation. Next, a mask of brain voxels was estimated to constrain later procedures. Images with artifacts were automatically identified as those frames on which total displacement relative to the previous frame exceeded 1 mm or where the average intensity within the brain mask deviated from the run mean by greater than three standard deviations using methods described in the Artifact Rejection Toolbox (*http://www.nitrc.org/projects/artifact_detect*). The functional data were spatially smoothed with a 6 mm FWHM kernel using the SUSAN algorithm from FSL, which restricts smoothing to voxels of similar intensity (Smith & Brady, 1997). Finally, the timeseries data were high-pass filtered by fitting and removing gaussian-weighted running lines with an effective cycle cutoff of 160 seconds.

Separately, the T1-weighted anatomical volume was processed using Freesurfer to segment the grey-white matter boundary and construct tessellated meshes representing the cortical surface (Dale et al., 1999). Functional data from each run were then registered to the anatomical volume with a 6 degree-of-freedom rigid alignment optimizing a boundary-based cost function (Greve & Fischl, 2009). The anatomical image was separately normalized to MNI152 space using FSL's nonlinear registration algorithm (Jenkinson, Beckmann, Behrens, Woolrich, & Smith, 2012). Following these steps, the linear functional-to-anatomical transformation matrix was combined with the nonlinear anatomical normalization parameters to derive a single transformation from native run space to MNI152 space.

Parametric Load Analysis

A linear model was fit separately to each functional timeseries using Gaussian least squares with local correction for temporal autocorrelation (Woolrich, Ripley, Brady, & Smith, 2001). A single task regressor with boxcar functions indicated the task load, with 0-back, 1-back, 2-back and 3-back being modeled with a -3, -1, 1, and 3 weighting, respectively, while a column of 1s modeled the main effect of task. Additionally, the instruction period before each block was modeled with a separate regressor. These regressors were then convolved with the canonical difference-of-gammas hemodynamic response function from the FSL software package (Jenkinson et al., 2012). In addition to the artifact indicator vectors described above, we included regressors for the six realignment parameters used during motion correction (i.e. translations along and rotations around the three main axes in native participant space) to account for residual noise variance. Following model fitting, the contrast effect-size and standard error images were normalized to group space using the transformation described above with tri-linear interpolation. Contrast effects were then combined across runs with a subject-level fixed effects model.

Mixed-effects higher-level analyses modeled the effect of the parametric regressor within the MOT and dual n-back group independently for pre- and post-training, while the longitudinal effect of training within each group and the group x time interaction was assessed with an ordinary least squares (random effects) general linear model using the post-pre parameter estimate maps as inputs. Correction for multiple comparisons was accomplished by first thresholding whole-brain maps at a z-score of 2.3, then cluster-correcting at $p < .05$.

Analysis of Independent Load Contrasts

This model was identical to the parametric model described above, but replaced the single parametric task regressor with four independent regressors, one for each of the n-back difficulty levels, with boxcar functions indicating the task blocks.

Regions/Networks of Interest Analysis

We focused analyses within the executive control network (ECN) and dorsal attention network (DAN) regions of interest (ROIs) derived from a population atlas of task-independent cortical networks (Yeo et al., 2011). As this atlas is defined in Freesurfer's common surface space, region labels were first warped back to the individual subject surfaces by inverting the spherical normalization parameters obtained during cortical reconstruction. Vertex coordinates within each of these labels were then transformed into the native functional space by inverting the linear functional-to- anatomical transformation for the first run. Finally, voxels were identified for inclusion within each region's ROI mask by projecting half the distance of the cortical thickness at each vertex and labeling the intersected voxels. This method produced ROIs that reflected the underlying two-dimensional topology of the cortex and minimized the inclusion of voxels lying outside of gray matter.

Functional Connectivity Analyses

The result of dual n-back training on functional connectivity within task-positive regions was assessed in the following way: ROIs were derived from the passive control group so that they were defined independently of the active training groups.. Because the greatest training-related activation changes occurred in the 2-back condition (reported below), the passive control group's 2-back vs. baseline contrast was used to obtain functional connectivity seeds. The positive activations from this contrast were thresholded at $p = .001$, then cluster-corrected at $p = .05$. The resulting clusters were masked with an anatomical mask from the Harvard-Oxford atlas including the lateral prefrontal cortex and lateral parietal cortices, yielding 4 ROIs: one prefrontal ROI and one parietal ROI in both the right and the left hemisphere. Each of these four ROIs was then reverse-normalized into native space for each participant and intersected with the subject-specific DAN and ECN ROIs described above, yielding 12 ROIs in total per participant (4 whole cluster ROIs, 4 DAN ROIs and 4 ECN ROIs).

For each prefrontal ROI, a timecourse over the experiment was extracted (using the mean activity of all voxels within the ROI). Connectivity each dual n-back level was assessed independently. For a given n-back level, the portions of the timecourse corresponding to seconds 8-30 of each n-back block were preserved (the first 6 seconds were omitted to avoid artifactual correlations emerging from the rising hemodynamic response), while the remainder of the timecourse was set to 0.

A general linear model was then created for each ROI using its specific n-back timecourse as the regressor of interest, and including traditional block regressors for the other 3 conditions, as well as the outlier volumes and motion parameter nuisance regressors described above. Finally, 6 additional nuisance regressors were included representing the first 6 principal components of the white matter and CSF timecourse, taken from a Freesurfer mask of white matter and CSF that had been eroded by 2 voxels to avoid contamination from partial volumes of gray matter voxels.

These models were then processed as described above through the fixed effects analysis for each subject, using FSL. Finally, for each of the 6 prefrontal ROI models, the mean parameter estimate measuring connectivity between that prefrontal ROI and each of the 6 parietal ROIs was extracted. These parameter estimates were included in 36 independent group x session mixed-effects ANOVAs for the dual 1-back, 2-back, and 3-back levels. The resulting 108 ANOVAs (3 n-back levels x 6 PFC ROIs x 6 parietal ROIs) characterize the effects of training on connectivity at each dual n-back level across the two training groups. To address the problem of multiple comparisons with this many analyses, the 36 p-values at each n-back level were adjusted using false discovery rate (FDR) corrections at a level of $p = .05$.

Connectivity/Behavioral Correlations

To assess the relationship between the connectivity changes described above and changes in performance on the in-scanner dual n-back task after 20 sessions of training, we calculated simple correlation coefficients between the change in connectivity metric and the change in behavioral performance. In order to avoid calculating this correlation for each of the 36 pairs of ROIs (and to avoid the concomitant problem of multiple comparisons), we calculated this correlation using only the whole-cluster PFC ROI to whole-cluster parietal ROI changes.

Results

Behavioral Results – Baseline performance

A one-way ANOVA across loads at baseline revealed that increasing dual n-back loads resulted in significantly decreasing accuracy (measured as hit rate – false alarm rate) [$F(3,114) = 108.3$, $p < .001$] and increasing reaction time [$F(3,114) = 125.1$, $p < .001$] (Figure 1). There were no significant differences at any loads between the two groups on either accuracy (p's > .11) or reaction time (p's > .14).

Behavioral Results – Changes in performance after training

A mixed-design ANOVA examining session x n-back load x group for accuracy revealed that there were significant main effects of load [$F(3,137) = 85.9$, $p < .001$], session [$F(1,37) = 30.9$, $p < .001$], and group [$F(1,37) = 5.84$, $p = .021$], as well as significant interactions of group x load

[F(3,111) = 6.22, p < .001], group x session [F(1,37) = 11.8, p = .002], load x session [F(3,111) = 24.4, p < .001], and group x load x session [F(3,111) = 21.3, p < .0001]. The main effects of this ANOVA showed increased accuracy with lower loads, increased accuracy post-training, and increased accuracy within the n-back group. The group x session interaction revealed higher accuracies a disproportionate gain in accuracy for the n-back relative to the MOT group after training.

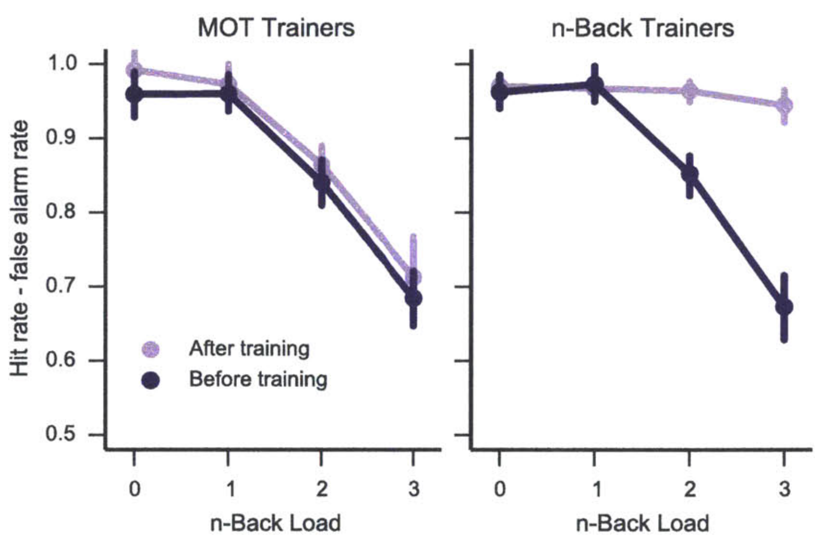
In order to better understand the interactions, session x group ANOVAs were performed separately for each load. Separate mixed-design ANOVAs at each dual n-back load show that the omnibus interaction was driven by significant group x session interactions only at the dual 2- and 3-back loads, where the n-back training group became both more accurate and responded more quickly than did the MOT training group. [0-back F(1,37) = 2.2, p = .14; 1-back F(1,37) = 1.6, p = .21; 2-back F(1,37) = 5.8, p = .02; 3-back F(1,37) = 28.7, p < .001] (Figure 1a and 1b).

Within-group repeated measures t-tests further confirmed that the 2-back and 3-back interactions were driven by the improved post-training accuracy within the n-back training group while the MOT training group did not significantly improve.

An identical mixed-design ANOVA examined the effects of session x load x group on reaction time. Results showed significant main effects of load [F(3,137) = 101.8, p < .001] and session [F(1,37) = 44.3, p < .001] but no main effect of group [F(1,37) = 0.02, p = .88]. However, there were significant interactions of group x load [F(3,111) = 5.33, p = .002], group x session [F(1,37) = 21.3, p < .001], load x session [F(3,111) = 30.7, p < .001], and group x load x session [F(3,111) = 20.4, p < .0001]. The main effects of this ANOVA showed faster reaction times with lower loads and after-training, but no difference between groups when collapsing across times. The group x session interaction, however, revealed faster reaction times within the n-back trainees in the post-training session.

Mixed-effects group x session ANOVAs run at each n-back load independently show the group x session x load interaction was driven by changes in reaction time at the dual 2-back and 3-back loads [0-back F(1,37) = 1.1, p = .30; 1-back F(1,37) = 1.15, p = .29; 2-back F(1,37) = 16.7, p < .001; 3-back F(1,37) = 48.8, p < .001]. (Figure 1c and 1d)

Within-group repeated measures t-tests further confirmed that the 2-back and 3-back interactions were driven by the improved post-training reaction times within the n-back training group while the MOT training group did not significantly improve.



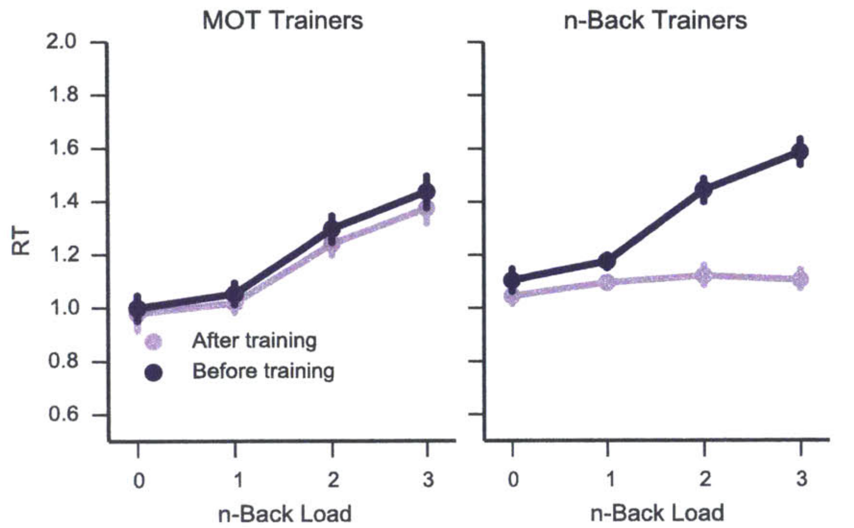


Figure 1: a) Sensitivity index of MOT training group before and after MOT training, across dual n-back load. b) Sensitivity index of dual n-back training group before and after training. c) Reaction times for responses to matching dual n-back trials in the MOT training group before and after training. d) Reaction times for the dual n-back training group before and after training. Error bars represent 95% within-subject confidence intervals.

Parametric Task Regressor – Activation Pre-training

A parametric model with weights -3, -1, 1, 3 respectively for the 0-, 1-, 2-, and 3-back blocks across both training groups during the pre-training dual n-back session revealed two extensive clusters of positive activation: the anterior cluster included bilateral superior, middle, and inferior frontal gyri, insular cortex, and striatum, with the peak voxel located inside insular cortex. The posterior cluster included the lateral occipital complex, precuneus and supramarginal gyrus (Figure 2), with the peak voxel located on the supramarginal gyrus. Peaks within those clusters were obtained from FSL's cluster utility using a minimum between-peak distance of 30 mm, and the regions associated with those peak coordinates were determined from the Harvard/Oxford atlas distributed with FSL. (Table 2).

Cluster	Num Voxels	z-Value	x	y	z	Region
Anterior	40996	9.76	32	22	2	Insular Ctx
		9.59	-32	24	2	Insular Ctx
		8.81	30	10	60	MFG
		8.78	0	14	54	SFG
		8.69	-46	24	30	MFG
		8.61	44	34	28	MFG
Posterior	15017	10.00	44	-42	48	Supramarg G, post
		9.76	-38	-48	42	Supramarg G, post
		8.53	-10	-68	60	Lat Occ Ctx, sup
		8.16	32	-70	52	Lat Occ Ctx, sup
		3.28	20	-56	24	Precuneous Ctx
		2.42	-30	-84	6	Lat Occ Ctx, inf

Table 2: Activation peaks within the two clusters identified with a parametric load regressor encoding 0-back through 3-back dual n-back blocks.

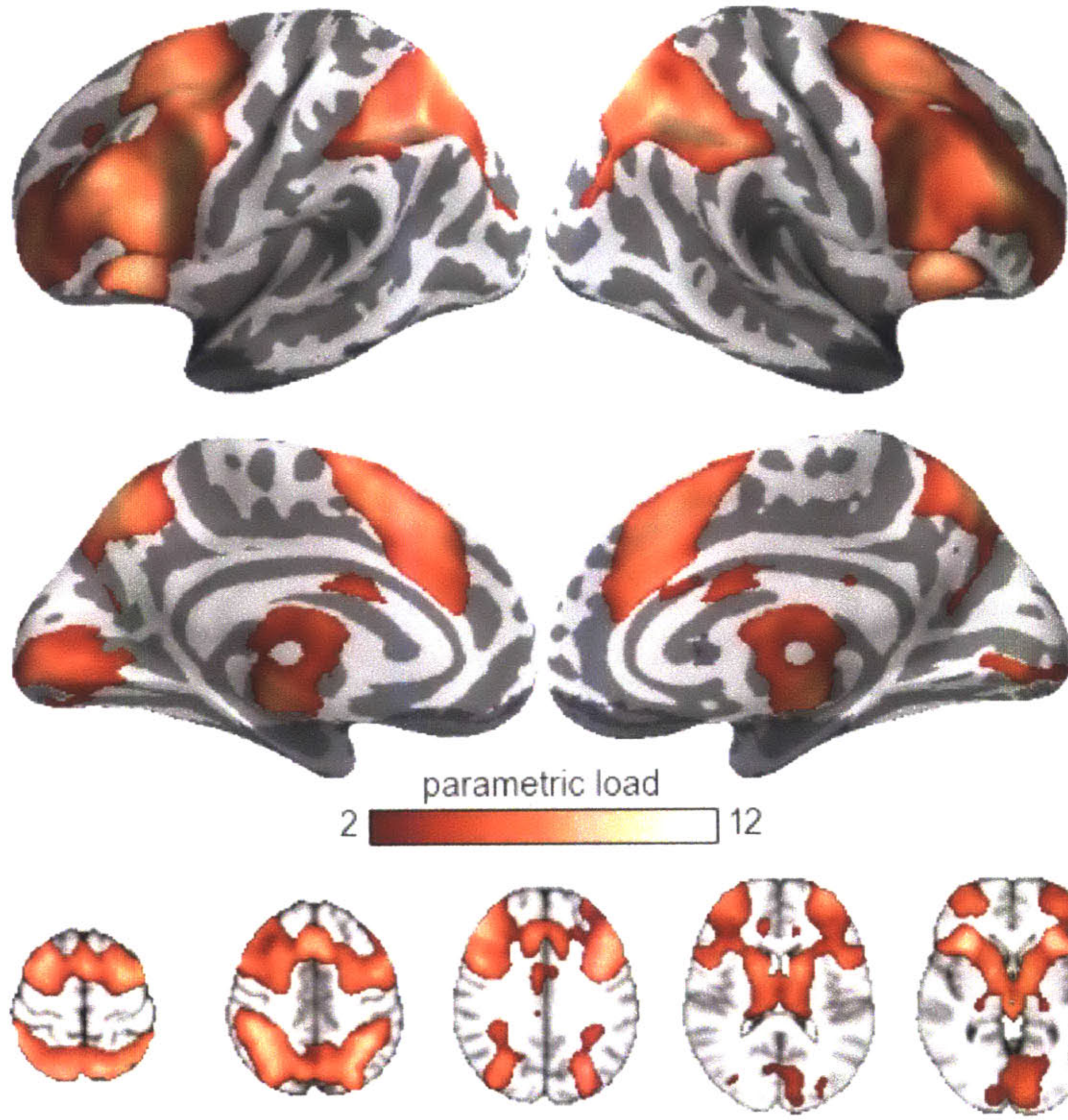


Figure 2: Pre-training functional activations: Increasing dual n-back load from 0-back through 3-back evoked functional activations in large regions of cortex, as well as in subcortical structures including the basal ganglia and thalamus. Load is modeled as a parametric regressor. Both training groups are included (n = 39). Corrections for multiple comparisons performed at a voxel threshold of $z > 2.3$, $p < .01$; cluster threshold $p < .05$.

Parametric Task Regressor – changes with training

The differential effect of training type on dual n-back activations was evaluated with functional maps showing the group x time statistical interaction over the load-based parametric regressor (effectively creating a group x time x load model). Regions with a significant interaction included the frontal poles, inferior, superior, and middle frontal gyri, insular cortex, supramarginal gyrus, angular gyrus and cingulate gyrus. (Figure 3) These interactions result from relatively reduced activations in the dual n-back training group after training, compared to similar activations before and after training in the MOT group. (In fact, the comparison of pre- to post-training activations within the MOT group reveals no clusters survive thresholding.)

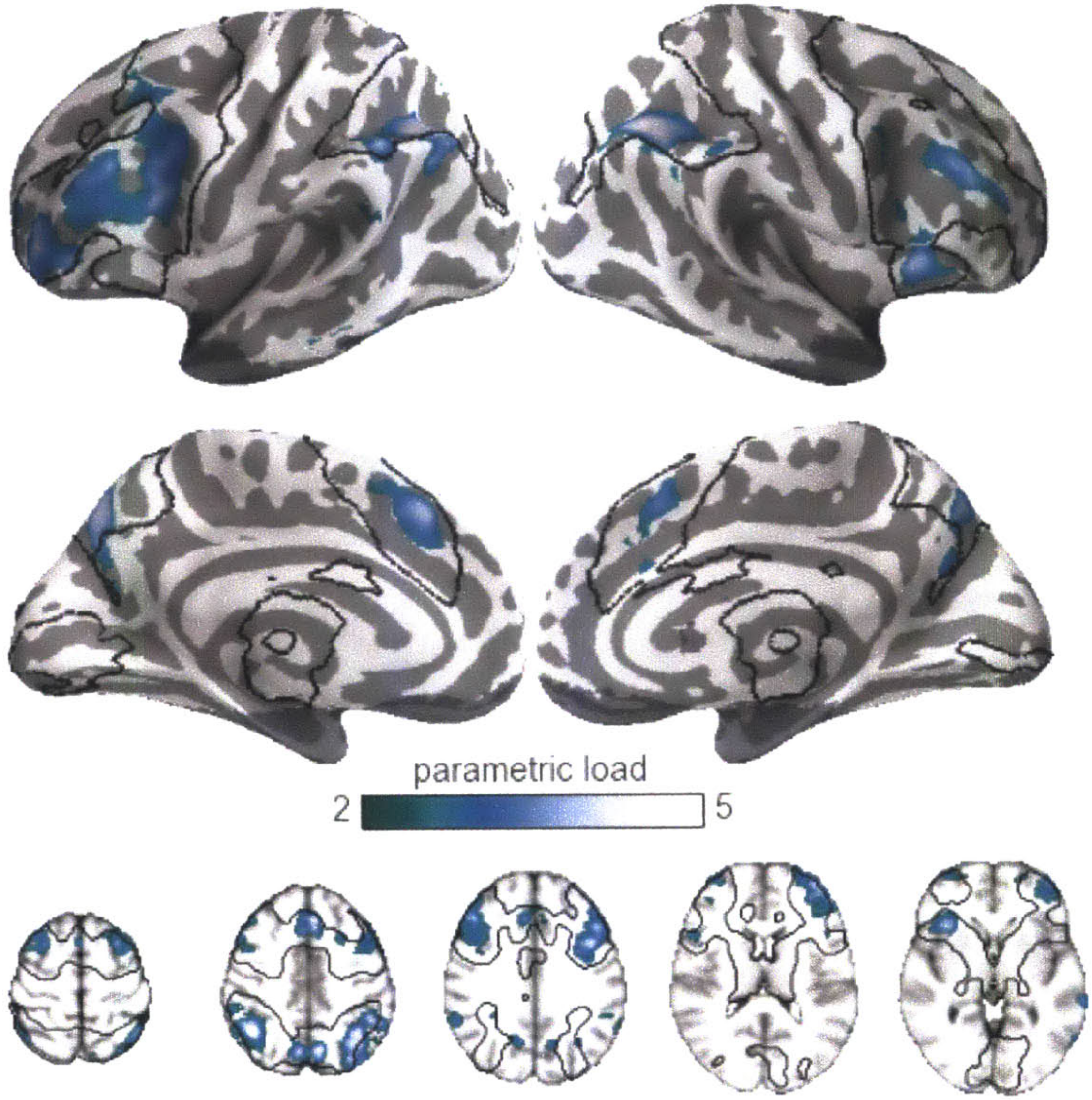


Figure 3: Regions displaying a significant group x time interaction with a parametric dual n-back task regressor. Black outlines indicate clusters of significant positive activation pre-training (see figure 2). Corrections for multiple comparisons performed at a voxel threshold of $z > 2.3$, $p < .01$; cluster threshold $p < .05$.

Individual task regressors – changes with training

The parametric regressor group x time interaction displayed regions with reduced activation after training in the n-back group. To clarify whether that reduction was across task levels or observed within a smaller set of levels, we again analyze the data with each n-back level having its own regressor. The results of this analysis reveal the reduced activation patterns observed in the parametric regressor were driven pre-dominantly by changes at the 2-back difficulty level, with additional significant posterior reductions in the dual 3-back (Figure 4).

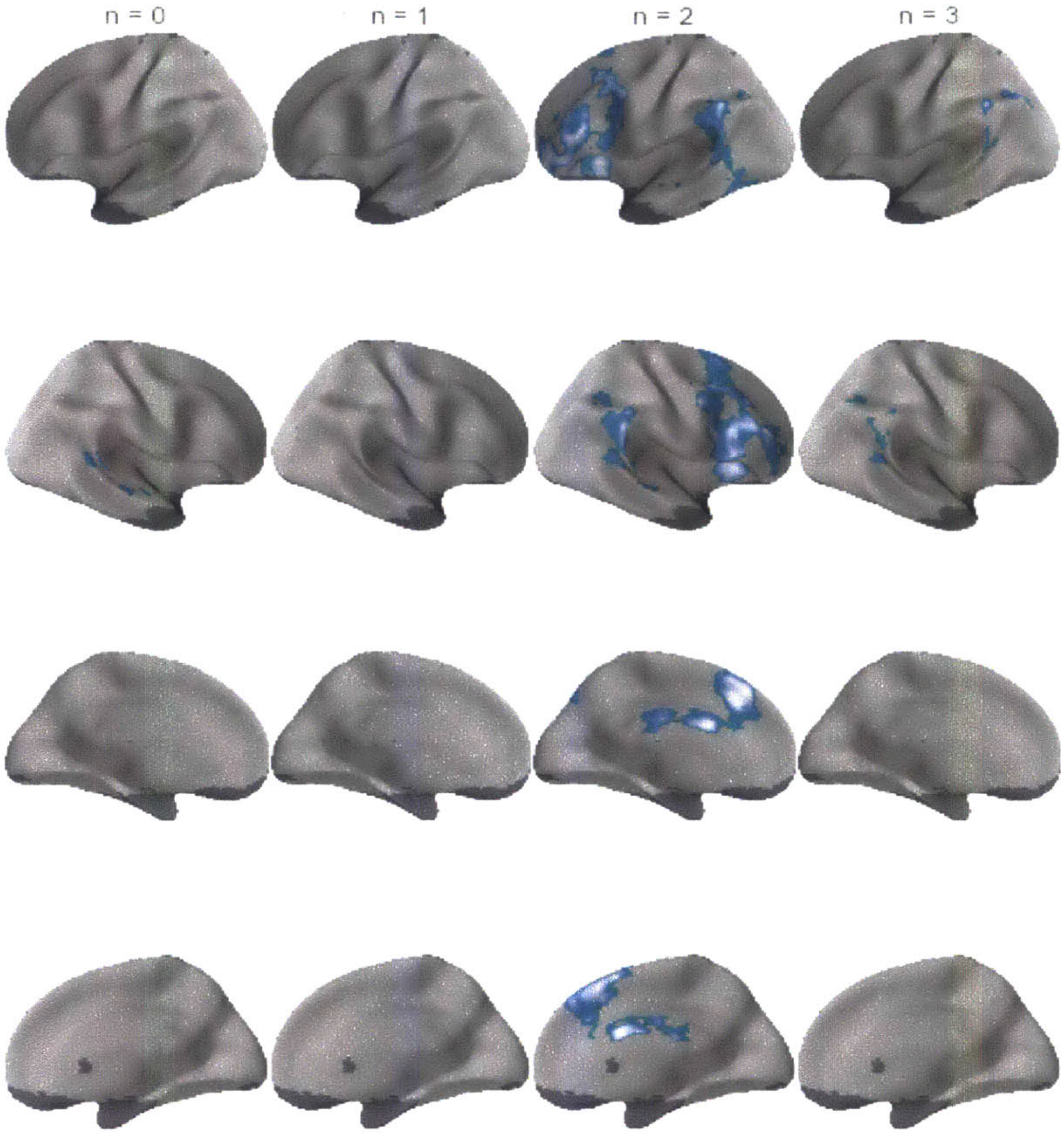


Figure 4: Clusters showing significant Group x Time interactions at each load of the dual n-back task. Corrections for multiple comparisons performed at a voxel threshold of $z > 2.3$, $p < .01$; cluster threshold $p < .05$

Executive Control and Dorsal Attention Network Regions of Interest

The executive control and dorsal attention networks, derived from resting-state functional connectivity analyses (Yeo, 2011), were examined separately in relation to training-induced changes in activation. The two networks were superimposed onto the pre-training activations from the independent no-contact control group using the parametric load regressor. (Figure 5) With the two exceptions of the occipto-temporal portion of the dorsal attention network and portions of the cingulate cortex within the executive control network, the task-evoked activations were largely overlapping with the two networks.

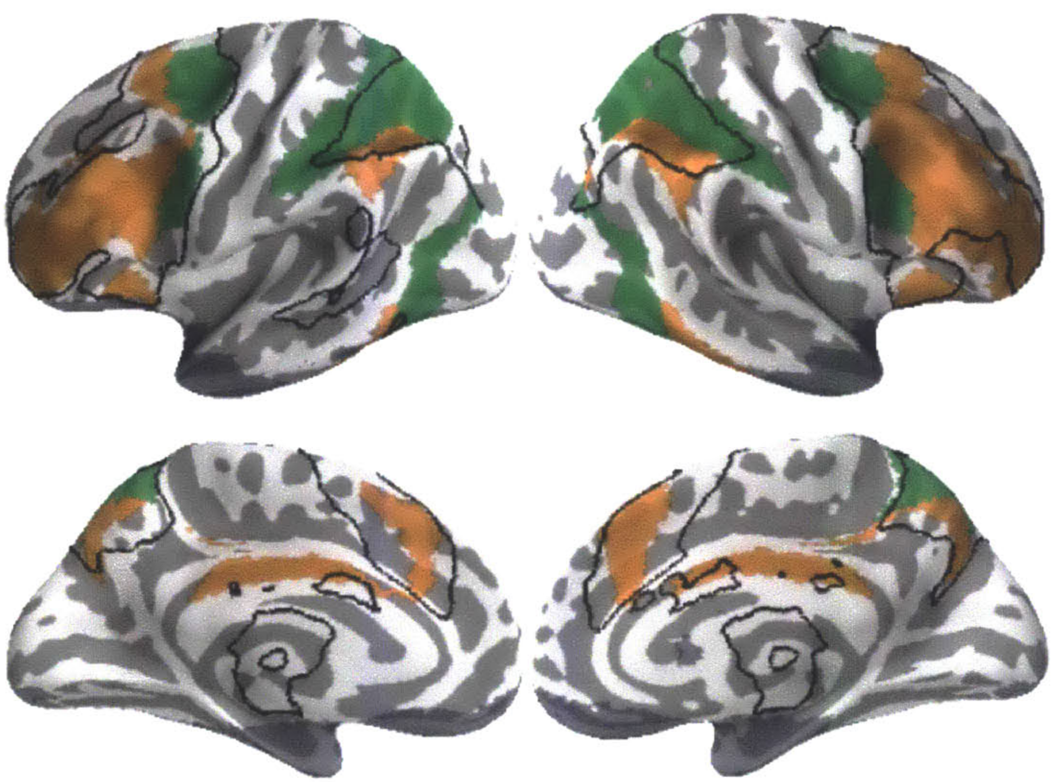


Figure 5: Independently defined dorsal attention network (green) and frontoparietal network

(orange) from Yeo, 2011. Significant positive clusters associated with the parametric load regressor are outlined in black.

A load x session x group x ROI mixed-effect ANOVA [F(3,111) = 9.14, p = .003] demonstrated a significant difference between activation changes in the two networks over time (Figure 6). Both the executive control and dorsal attention networks exhibited load-dependent activations before training. After n-back training, there was reduced activation in the executive control network for the demanding 2-back and 3-back conditions, but no reduction in activation for any condition in the dorsal attention network. MOT training did not alter activation between the pre- and post-training sessions.

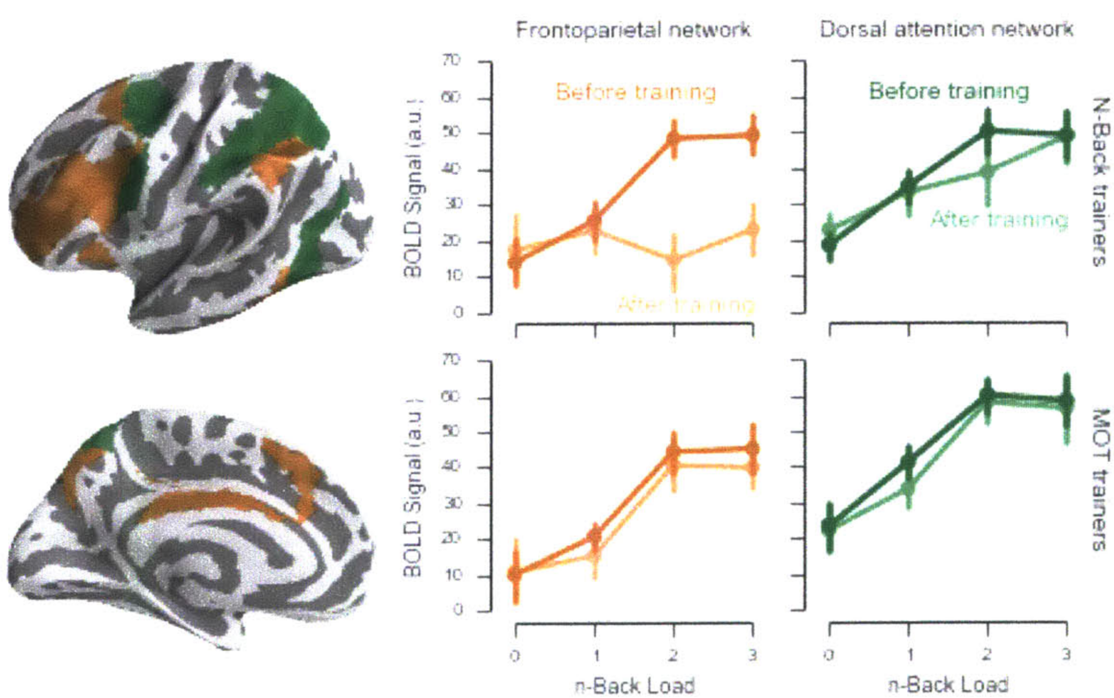


Figure 6:

Activation levels within executive control and dorsal attention networks at each level of the dual n-back task after either dual n-back or MOT training. After training, the dual n-back group shows reduced activation in the executive control network but not in the dorsal attention network.

Frontal-parietal connectivity – changes with training

We examined training-related changes in functional connectivity (temporal correlations) between prefrontal and parietal ROIs defined from an independent passive control group during the dual 2-back task, we assessed training-related changes in functional connectivity.

Figure 8 displays the session x group interaction scores from mixed effects ANOVAs assessing changing in connectivity between each pair of prefrontal and parietal ROIs in the dual 1, 2, and 3-back conditions. In the dual 2-back condition, each of the 36 anterior-posterior pairings revealed a significantly greater connectivity increase in the dual n-back training group than in the MOT group (all FDR-adjusted p-values $< .01$). However, unlike the reduced activation observed after dual n-back training, the increased connectivity did not adhere to the frontoparietal and dorsal attention network resting state boundaries. In the dual 1-back condition and dual 3-back conditions, no significant group x session interactions showing changes in functional connectivity survived corrections for multiple comparisons (all FDR-adjusted p-values $> .8$ in the dual 1-back condition, all FDR-adjusted p-values $> .4$ in the dual 3-back condition).

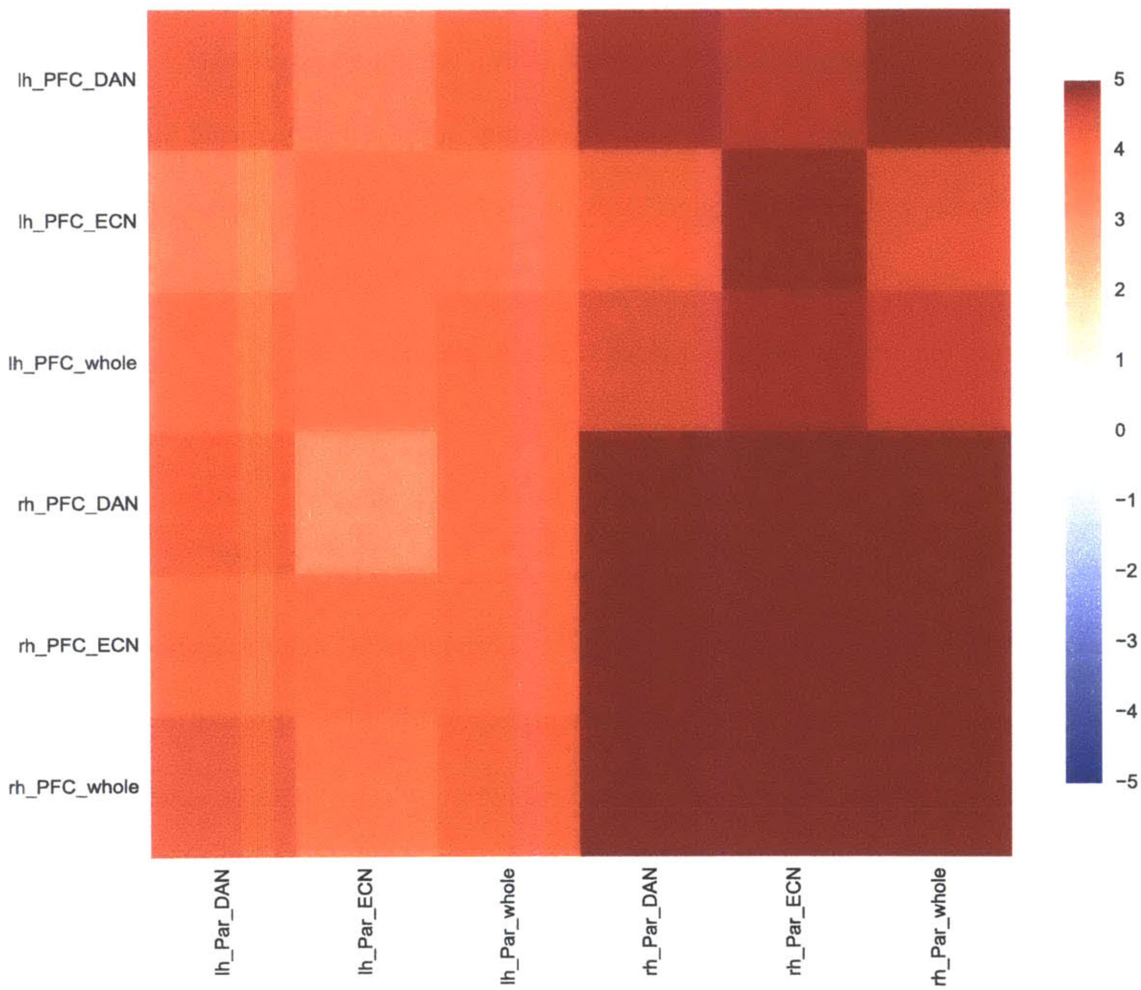


Figure 8:

Heatmap showing t-values for the group x session connectivity increases using ROIs independently defined in the passive control group. For each prefrontal – parietal pairing, functional connectivity measures significantly increased in the dual n-back training group compared to the MOT training group (all FDR-adjusted p-values < .01).

Correlations between connectivity changes and in-scanner performance

Simple correlations describe the degree to which the behavioral improvements within the n-back group were predicted by the functional connectivity changes described above during the dual 2-back condition. Accuracy improvements in the dual 2-back scanner task were predicted by the magnitude of the frontoparietal connectivity increase within the left hemisphere ($r = .57$, $p = .01$) and within the right hemisphere ($r = .45$, $p = .05$). Accuracy improvements during the dual 3-back condition were only predicted by connectivity increases in the left hemisphere ($r = .49$, $p = .03$), while the right hemisphere was not predictive ($r = .27$, $p = .25$). Connectivity increases did not predict changes in reaction times (all r-values $< .22$, all p-values $> .35$).

Discussion

Functional brain plasticity associated with a large increase in dual n-back WMC was characterized as highly specific in multiple ways. First, such plasticity occurred selectively in the dual n-back training group, who more than doubled their WMC (Thompson et al., 2013), and not in the MOT training group, who showed neither behavioral improvement nor plasticity associated with the dual n-back task. Second, the n-back training group showed both behavioral gains and reduced activation selectively for the WM-demanding 2-back and 3-back conditions. Third, although both the executive control and dorsal attention frontoparietal networks demonstrated load-dependent activation, only the executive control network exhibited reduced activation with training. Fourth, the ECN and DAN both exhibited increased functional connectivity accompanying expanded WMC, and these reductions in functional connectivity correlated with behavioral improvement in the 2-back condition.

Pre-training activation in dual n-back task

The pattern of load-dependent activation at pre-training was generally consistent with prior neuroimaging studies (Owen et al., 2005), which have typically reported activations in lateral premotor cortex; dorsal cingulate and medial premotor cortex; dorsolateral and ventrolateral prefrontal cortex; frontal poles; and medial and lateral posterior parietal cortex. Specific patterns of activation vary in relation to stimulus and task dimensions, and only three studies have examined the functional activations associated with a simultaneous visual-spatial and auditory-verbal dual-back task (Buschkuehl, Hernandez-Garcia, Jaeggi, Bernard, & Jonides, 2014; Jaeggi et al., 2007; Yoo, Paralkar, & Panych, 2004). The pattern of load-dependent activation observed here, including frontal, parietal, temporal, and subcortical activations, resembles those reported in the prior dual n-back studies.

Training-related activations in dual n-back task

Functional brain changes mirrored behavioral changes in performance after training. The MOT group exhibited no improvement in dual n-back performance despite great improvement on the MOT task (Thompson et al., 2013), and also exhibited no difference across sessions in functional activation. The dual n-back group showed no gain in performance for the 0-back and 1-back conditions, presumably performing at ceiling from the outset, and exhibited no difference across sessions in functional activation at those loads. The training did, however, substantially improve performance in the more demanding 2-back and 3-back conditions, which the n-back training group performed with the same ease (measured by accuracy and reaction time) as the 0-back and 1-back conditions. Correspondingly, there were significant and widespread reductions in activation, most notably in bilateral inferior and middle frontal gyri, insular cortex and intraparietal sulci.

Frontal and parietal regions were divided into two separable networks based on resting-state analyses of functional connectivity, namely an executive control network and a dorsal attention network (Yeo et al., 2011). Both networks responded with greater activation to greater load in the initial session. The training-induced reductions of activation occurred exclusively in the executive control network, including DLPFC and superior parietal cortex. Prior studies of training-induced changes in activation have not examined the two networks independently.

The dissociation between training-induced reductions in activation across the two networks must be related to the distinct roles of those networks. Although both networks are thought to play a role in exerting top-down control on cognitive processing, they have been associated with different functional roles. Regions within the executive control network are characterized by the flexibility of their representational content (Duncan, 2010) and ability to sustain distractor-resistant representations of information relevant for goal-directed processing (E. K. Miller, Erickson, & Desimone, 1996; Waskom, Kumaran, Gordon, Rissman, & Wagner, 2014). Dorsal attention regions, in contrast, appear to contain prioritized topographical maps of visual space (Silver & Kastner, 2009) that are used for internally directed visual attention (Corbetta & Shulman, 2002), a possible mechanism of visual working-memory maintenance (Sreenivasan, Curtis, & D'Esposito, 2014). Our results thus suggest that training reduces the demand on the higher levels of the hierarchical system that supports the maintenance and updating of active working-memory traces.

WM training studies have reported multiple patterns of functional plasticity associated with gains in WM performance including increased activations or decreased activations. The variety of changes in activation may reflect not only the variety of tasks, but also the duration of training (ranging from a single session of training to our study of a month of intensive training) and the initial and final levels of performance. In our study, participants were considerably above chance in all conditions at the outset, improved their performance dramatically on the more difficult conditions (which were far below their WM capacity from the month of intensive training), and exhibited large reductions of activation in the more demanding conditions.

Training-related changes in functional connectivity

Working memory training also increased functional connectivity between frontal and parietal regions during task performance in the scanner during the dual 2-back and dual 3-back conditions (the two conditions that also exhibited behavioral improvement and reduced activations). In contrast to the post-training decreased activations, these changes in functional connectivity occurred in both the executive control and dorsal attention networks. Most importantly, the degree to which connectivity increased significantly correlated with the amount of improvement observed in dual 2-back and 3-back accuracy. Such a brain-behavior correlation was not observed for the reduction in activation. Although some prior studies of WM training have also reported the functional activation decreases we observe, to our knowledge no prior studies of WM training have reported changes in functional connectivity during WM performance as a consequence of training, nor have they reported links between neuroimaging measures and behavioral performance on the trained task.

Training-related changes and the neuroimaging of learning

When considering the functional brain effects of learning, there are three primary changes that might be observed: activation *decreases*, activation *increases*, or activation *shifts* (see Kelly, Foxe, & Garavan, 2006 for a detailed discussion of this topic). In the case of decreasing activation, the network of brain region initially supporting task performance remains the same, but activation in some or all of those regions decreases as performance improves. This reduced activation in task-relevant regions is often interpreted as evidence for increased efficiency in the neural processing relevant to the task (but see Poldrack, 2014). For example, learning might result in sharper, sparser representations of encoded stimuli, meaning that a reduced neural population is necessary to represent task-relevant information; or learning may reduce the number of neural circuits being recruited for the task, resulting in less observed metabolic activity (Haier et al., 1992). An alternative explanation is that learning could decrease the time that participants spend on the task, which could result in a similar reduction of the BOLD response (Poldrack, 2000).

A second possibility is that, after learning, the network of task-relevant brain regions remains the same, but that activation in these regions *increases* with improved behavioral performance. Increased brain activation is commonly observed as a consequence of extensive practice on motor tasks (e.g., string playing [Elbert, Pantev, Wienbruch, Rockstroh, & Taub, 1995] or sequential finger movements [Ungerleider, Doyon, & Karni, 2002]). Increases in activation accompanying improved performance have been argued to reflect an increase in the spatial extent of activation (a larger population of cells) due to increases in the size of cortical representations (Pascual-Leone, Amedi, Fregni, & Merabet, 2005) and a stronger neural response in existing areas (Kelly et al., 2006).

The third possibility is that learning reflects a shift from one set of brain regions to another, such that some regions show decreases in activation over the course of learning, while others show increased activation, conceptualized as either a "reorganization" or "redistribution" of activation. This is frequently observed during the learning phases of a sensory or motor task, where a common finding is that prefrontal cortical activation decreases while more domain-specific cortical areas show increased activation (e.g., Sakai et al., 1998; Shadmehr & Holcomb, 1997), This pattern may reflect a 'scaffolding-storage' framework, where domain-general attentional and control regions (prefrontal cortex, anterior cingulate cortex and posterior parietal cortex) perform the 'scaffolding' role, supporting demands during the learning of a new, effortful task. This 'scaffold' then falls away as domain-specific sensory and motor 'storage' regions are able to support task performance on their own (Petersen et al., 1998). In other cases, there could be a shift between the natures of representations involved, such as moving from a visuospatial to a linguistic basis of performance and therefore changing the hemispheric basis of early, naive versus later skilled performance (Poldrack, Desmond, Glover, & Gabrieli, 1998).

Neither the scaffold-storage hypothesis nor the representational shifts make clear predictions for WMC training-induced neuroplasticity, as the "scaffolding" areas are the same areas required for successful performance of a working memory task and it seems unlikely that the representations of simple letter stimuli will dramatically change. This has led to speculation that, unlike patterns seen in sensory and motor learning, prolonged WM training would result in increased neural activations corresponding to increased working memory capacity (Buschkuehl et al., 2014). This research contradicts that hypothesis in favor of a more general "neural efficiency" hypothesis across domains..

Conclusions

Interest in the dual n-back task arises from at least two sources. Scientifically, the task is a complex working memory task that exercises each of the putative constructs in the suite of "executive functions" (Miyake et al., 2000) – monitoring and maintenance in the encoding of incoming stimuli, inhibition in the avoidance of lure trials, and switching in the requirement of encoding stimuli from two domains simultaneously. Beyond the purely academic appeal, however, is the prospect that training on the dual n-back task could provide transfer to other tasks (Au et al., 2014; Jaeggi et al., 2008), based on the idea that plasticity within a common neural substrate could improve all tasks depending upon that substrate (Dahlin et al., 2008; Jonides, 2004). As shown here, the cortical regions recruited by a dual n-back task are nearly identical to those described as supporting human intelligence (Jung & Haier, 2007), yet in this group of participants, no improvement was seen on any other cognitive assessment despite behavioral improvements on the trained task (Thompson et al., 2013) or the robust functional changes on neuroimaging measures reported above.

Although training on a dual n-back task did not yield far transfer within our population, it did enable remarkable performance on the trained task itself, with some participants becoming able to perform a dual 9-back after 20 sessions of training. Confirming previous reports, this feat was accompanied by reduced activation during performance of the originally challenging dual 2 and 3-back conditions. Additionally, we report that the reduced activations were not present across the entirety of the frontal and parietal task-responsive regions, but were confined to the executive control network in support of the "neural efficiency" hypothesis. Finally, it was the increased frontoparietal connectivity that accompanied the reduced functional activations that was most predictive of behavioral improvements across individual subjects.

Chapter 4: Adaptive Training of Attentional Tracking Improves Both Speed and Capacity

Introduction

Research on human capacities for attention and working memory has proceeded at a break-neck pace since George Miller's classic "The Magic Number 7 +/- 2" paper was published in 1956 (G. A. Miller, 1956). Nearly 60 years later, we have a far more solid understanding of what the behavioral limits of capacity actually are (particularly the limits of healthy North American undergraduates), but uncertainty surrounds the origins and properties of those limits. In the multiple object tracking (MOT) task, for example, two behavioral limits can be easily estimated in any participant – the speed at which a set of targets can be successfully tracked and the number of targets that can be tracked at a given speed. Several questions surround these two limits, however. Are they fixed, perhaps reflecting an immutable low-level property of the visual system, or are they plastic? Do these two attentional limits emerge from a common underlying source, or are they independent? These questions have proven difficult to study with traditional analyses (both behavioral and neuroimaging), but are perhaps more easily studied through training studies. Here, we used an adaptive training program to train participants on the MOT task for 20 days, and assessed the extent to which training could improve tracking speed and the extent to which speed training transferred to capacity limits, under the assumption that if speed and capacity limits are strictly independent, gains in one should not influence the other.

Multiple Object Tracking Overview

One of the most frequently used experimental tools with which to study human visual attention is the multiple object tracking task (Pylyshyn & Storm, 1988). In these experiments, participants are presented with a display of identical objects. Some of those objects are identified as targets to be subsequently tracked, after which the entire set of objects begins moving. When the movement stops, the participant is asked to identify the original target objects.

Can MOT performance be improved?

A fundamental question is whether MOT speeds or the number of items tracked can be enhanced through training. Many aspects of our visual attention system have proven relatively inflexible in the face of training attempts (e.g., change blindness does not improve for untrained stimuli [Gaspar, Neider, Simons, McCarley, & Kramer, 2013]), while others are more tractable (e.g., training improves perception of Gabor orientation, [Lu, Chu, Dosher, & Lee, 2005]). While MOT has been often used to identify individual differences in attentional performance (Drew & Vogel, 2008; Oksama & Hyönä, 2010), it is unknown whether those differences are stable and rigid (perhaps arising from low-level cortical features dictating the resolution of attentional spotlights) or if they are amenable to improvement.

Some previous research suggests that that practice on the task yields modest improvements in MOT abilities. For example, basic task familiarity and adoption of learned strategies leads to a modest improvement in tracking ability (Yantis, 1992). The evidence for training-dependent improvement is less clear: Previous research programs using MOT as an outcome measure for other types of training have either shown modest benefits (Green & Bavelier, 2006) or no benefits at all (Boot, Kramer, Simons, Fabiani, & Gratton, 2008; Schwab

& Memmert, 2012). The one group that has attempted to train MOT directly (Legault, Allard, & Faubert, 2013) in healthy young adults adaptively trained participants on a 3-target version of a 3-dimensional MOT task for 5 30-minute sessions spread over 5 weeks, and reported modest improvements in tracking speed (approximately 40%) that appeared to plateau after the first few sessions, suggesting that tracking speed was actually relatively resistant to extended training.

A common cause for speed limits and number limits?

Behavioral data suggest that speed limits and number limits are not independent: the speed limit on tracking depends on the number of targets, and number limit depends on their speed (Alvarez & Franconeri, 2007; Holcombe et al., 2014) However, neuroimaging data suggest speed limits and number limits may ultimately be caused by separate underlying neural mechanisms (Shim et al., 2010). More specifically, activation levels in the posterior parietal cortices reflect changes in capacity while remaining insensitive to changes in speed. The frontal eye fields, in comparison, respond differentially to changes in both speed and quantity of targets. One implication from this finding is that the speed limits and capacity limits may reflect separate and independent limits on attentional tracking. This observation makes predictions on the transfer from speed-training to trackable items. If item capacity is limited by attentional pointers residing in the speed-insensitive parietal cortices, speed-training should not provide benefits to capacity.

The current study: training on speed, testing transfer to number

In this research, we estimated how fast participants could track 4 objects (pre-training baseline). We then examined whether training the speed at which 4 objects can be tracked would also improve the number of objects that could be tracked at the pre-training baseline speed. Our results demonstrate two novel findings: first, it is possible to train tracking speed to an extent we found surprising: after twenty days of tracking four objects, individual tracking speeds increased an average of 240%. Secondly, gains in tracking speed transferred to capacity. When asked to track varying numbers of targets using a constant tracking speed matched to their pre-training baseline, participants were able to track more items successfully, even though their training was based exclusively on increasing tracking speed. These results stand in contrast to a control group who trained on an adaptive dual n-back training task, involving a spatial working memory component, and showed no significant gains in either tracking speed or attentional tracking capacity. Taken together, these results demonstrate that neither tracking-speed nor tracking-quantity limits are fixed, and suggest that speed-limits and number-limits reflect shared underlying constraints on attentional tracking.

Methods

Overall experiment design

The data reported here is a subset from a larger study on the cognitive and neural effects of cognitive enhancement training. As part of this larger study, participants completed a pre-training assessment that included approximately six hours of behavioral testing spread across three days, and two hours of structural and functional magnetic resonance imaging (imaging data reported separately; other behavioral testing procedures and results reported in (Thompson et al., 2013). After this pre-training session, participants completed twenty sessions of adaptive training in the lab, followed by an immediate post-training assessment. A subset of participants also completed a six-month post-training assessment.

Participants were randomly assigned to receive training on a multiple-object tracking task (MOT), or a dual n-back working memory task. The details of the dual n-back training are reported elsewhere (Thompson et al., 2013), and the current report focuses on the MOT training group. However, for the purposes of comparison, we report the performance of both groups on the MOT tasks here.

Participants

Recruitment. Participants were recruited through web advertisements, physical flyers, and e-mail sent to the Northeastern and Tufts mailing lists. Participants were required to be adults between the ages of 18 and 45, right-handed, in good health, and not taking any drugs. All participants provided informed, written consent before participation. This study was approved by the Massachusetts Institute of Technology Institutional Review Board (PI: Leigh Firn).

Payment. Participants were paid $20 per training session, with a $20 bonus per week for completing all five training sessions in that week. All participants were paid $20 per hour for behavioral testing, and $30 per hour for imaging sessions.

Retention. In addition to the weekly bonus payment for completing all five sessions in that week, participants were emailed on a weekly basis congratulating them on their attendance, alerting them of their bonus, and informing them of the progress they had made in training that week. This email was intended to be motivational, so the email highlighted new achievements from the previous week (e.g., a new peak in a performance measure). All 39 participants completed the entire study.

Exclusions and deviations from protocol. There were 19 participants assigned to the MOT training group, and 20 assigned to the dual n-back control group. Through a combination of experimenter and technical errors, 4 participants from the MOT group and 2 participants from the dual n-back training group had their tracking capacity measured at a different speed after training than before training. For this reason, these participants have been excluded from further analysis. Of the remaining 15 MOT participants and 18 n-back participants, 3 from each group had their initial speed assessment measured with 40 trials rather than 90, based on an earlier version of the experimental protocol, and 1 participant from each group inadvertently pressed the "quit" key during the speed assessment, limiting the number of speed trials to 60 and 68 trials.

Status	N (female)	Mean Age (STD)	Mean Grade Level (STD)	Mean IQ (STD)
MOT	15 (10)	21.5 (2.2)	15.8 (1.6)	120.5 (7.2)
N-Back	18 (12)	21.0 (2.0)	15.5 (1.3)	118.9 (9.5)

Table 1: Participant demographics. IQ is calculated from the full-scale IQ score on the Weschler Abbreviated Scale of Intelligence (CITE).

<u>*Apparatus*</u>

The multiple object tracking task was programmed in the Psychophysics Toolbox 3, (Brainard, 1997), and run on an IBM ThinkPad laptop with a 15.4" display using a resolution of 1280 x 1024. Participants were asked to maintain fixation at a comfortable distance from the screen.

Sixteen green circles (20-pixel diameter, approximately 1.25°) were presented on a black background (900 x 700 pixels). A gray fixation point ("+") was presented at the center of the display. The circles moved at a constant speed and were "repelled" by each edge of the display and by other items with decreasing strength over distance, such that the items "avoided" each other. The circles changed direction to avoid other items and were never closer than 80 pixels (approximately 4°, measured center to center) to another circle.

Multiple Object Tracking Speed-limit Assessment

To assess the maximum speed at which participants could reliably track moving objects, we followed the general procedures from (Alvarez & Franconeri, 2007). Participants were asked to track 4 dots among 12 distractor dots. At the beginning of each trial, 4 target dots were identified in green for 500ms while all dots remained stationary. For the next 2500ms, all 16 dots moved while the target dots remained identified in green. At that point, the 4 target dots turned black, and for the remaining 8500ms of the trial, the target dots appeared identical to the distractor dots while the participant attempted to keep track of which dots were targets. Finally, participants identified the 4 tracked dots using a mouse and were given feedback.

The initial speed at which items moved for each participant was determined by a self-assessment task in which participants used the cursor keys to make a continually moving display of targets move slower or faster and reported the speed at which they thought they could reliably track four targets. This was followed by a thresholding procedure over the following 90 trials in which the speed of the moving dots increased by .5 degrees of visual angle/second every time two trials in a row were answered correctly, and decreased by .5 degrees/second every time two trials in a row were answered incorrectly. To count as a correct trial, all 4 targets were required to be identified correctly. Participants were allowed to take breaks, as needed. The speed of the final trial was the dependent variable.

Multiple Object Tracking Number-limit Assessment

Following the speed assessment, participants completed a MOT task in which the speed remained constant while the number of targets varied from 3 to 8 objects. For both the pre-training and post-training sessions, the tracking speed was set to be the final tracking speed from the initial session's 4-object speed assessment. After 6 practice trials, participants completed 12 trials at each load, intermixed in a randomized and counter-balanced order. For each trial, an estimate was calculated for the number of targets tracked on that trial, based on the number of targets successfully identified and the likelihood of guessing a target correctly through chance. The equation used to estimate items tracked was the same as that used in (Alvarez & Franconeri, 2007; Hulleman, 2005), provided here:

$$P(\text{correct}) = [C + (n - C) * (n - C)/(m - C)]/n.$$

where P(correct) is the average proportion of targets accurately clicked, C is the number of targets actually tracked, n is the number of targets, and m is the total number of items in the display. One variation from calculations in the previous study was that this equation was calculated in an iterative fashion to avoid under-estimating capacity. Because the maximum estimable capacity for trials with 3 targets is 3, including those trials in an average for a participant who was capable of tracking more than 3 objects at the test speed would artificially dilute the capacity estimate. To remedy this, we first calculated estimated capacity using all trials, then re-calculated the capacity using only trials that had more targets than the original estimate. If this raised the estimate above the next integer, the process was repeated until a stable estimate was reached. Three outcome measures were derived from the capacity assessments: 1) the number of completely accurate trials at each load; 2) the number of accurately identified targets at each load; and, 3) an overall estimated capacity score (k).

Participants were asked to return approximately six months after the completion of training later to examine the status of their improvement on the trained tasks. (Average number of days before the follow-up testing was 190 days, with a minimum of 133 days and maximum of 252 days. This time was not significantly different between groups [$t(13) = .57$, $p > .5$]). Although some participants were unable to return for follow-up testing (primarily due to post-graduation dispersal), useable data from 6 MOT participants and 8 dual n-back participants were collected, while data from two follow-up subjects was excluded due to being collected at the incorrect tracking speed.

<u>*Training protocol*</u>

Pre-training assessment. Participants completed a number of pre-training tasks as part of a larger study on cognitive enhancement training (see Thompson et al., 2013 for details). Among these tasks were the MOT speed-limit assessment and number-limit assessment described above.

Training. Participants completed 20 training sessions in which they performed 90 adaptive tracking trials per day, as described above. (Due to experimenter error, three participants had some days of training with 60 trials instead of 90 trials. These days were during the first half of the training period, and no subject had more than three short days.) The initial speed of the tracked objects was determined by the final speed of the pre-training baseline MOT session, which was reached via the thresholding procedure described above. On subsequent days, the first trial's speed was set to the speed of the last trial on the previous training day.

Adaptive training sessions lasted approximately forty minutes per day, and participants were asked to commit to one training session per day, Monday through Friday, at a consistent time. In the event that a training session was missed, participants were allowed to train on the weekend, or to train twice in one day, so long as the two sessions were separated by at least three hours of time. This option was used by 3 of the participants (with at most 3 double-session days). Participants completed the 20 training sessions in an average of 28.7 days (min 21 days, max 42 days).

Immediate post-training assessment. After training was completed, post-training behavioral testing and imaging were administered as soon as possible. (Average number of days between last training session and post-training testing was 4.3 days, with a minimum of 0 days and maximum of 14 days). Two participants were tested on the final training day, with at least 3 hours between the last training session and the post-testing session; all other participants were tested at least a day after the last training session. Among the tasks completed were the speed-limit assessment and number-limit assessment. The number-limit was assessed at the participant's pre-training baseline speed.

Results:

Adaptive training increases tracking speed for four targets

In the MOT training condition, participants improved from an average tracking speed of 8.8 degrees/second (SD = 3.2) over the first three days of training to an average speed of 14.9 degrees/second (SD = 4.2) over the last three days. [$t(18) = 11.6$, $p < .0001$]. Although there was a range of improvement in this condition, all participants were able to track items at least 12 degrees/second at some point during their training, with six participants becoming able to track 4 targets moving at faster than 20 degrees/second. On average, participants were able to track items 143% faster after training than before training. **[Figure 1]**

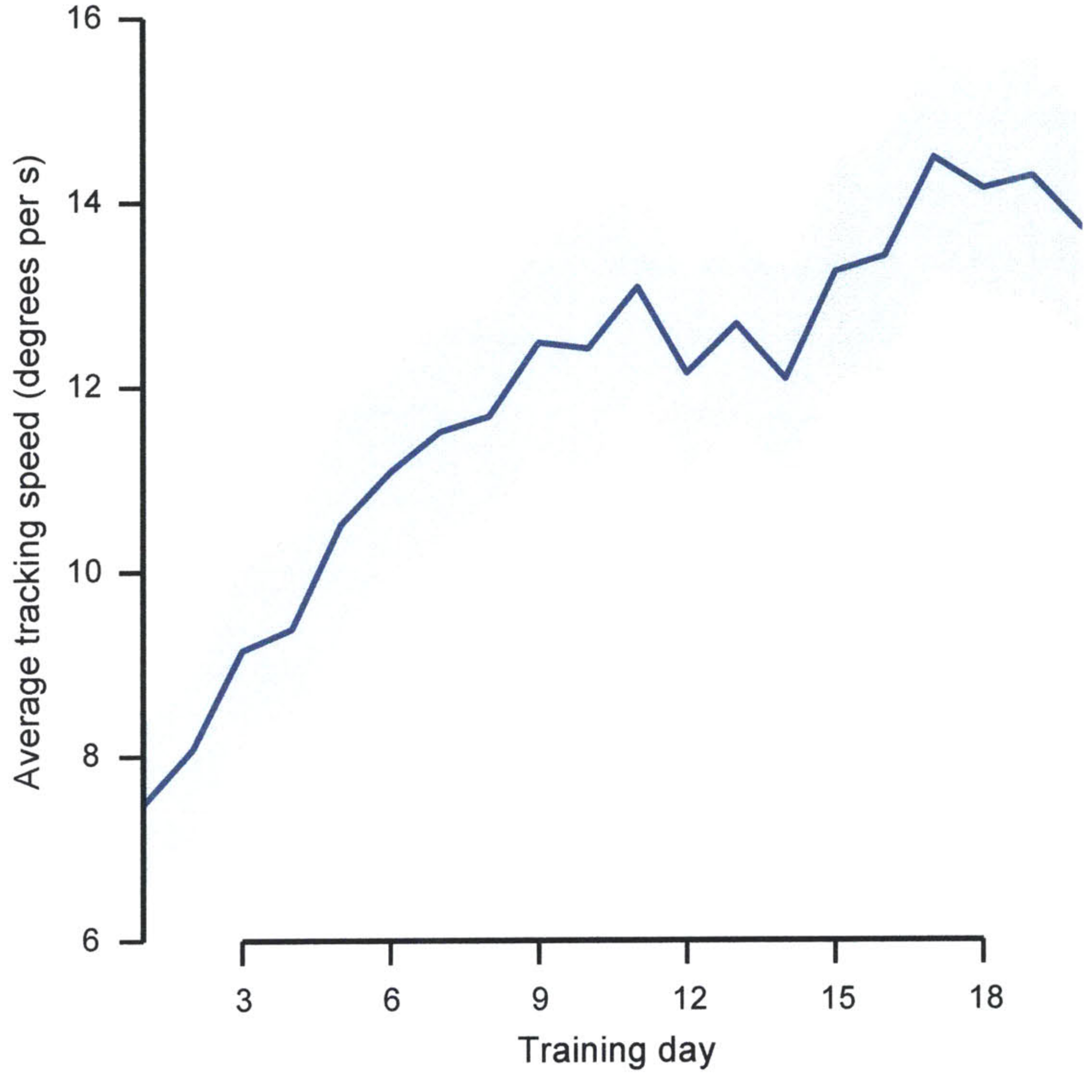


Figure 1: Mean object tracking speed at the end of each day of MOT training. Shaded area represents bootstrapped estimate of standard error of the mean.

Improvements on both the n-back and MOT tasks were specific to their training group. Comparing performance on these two tasks during the behavioral testing before and after training reveals a double-dissociation between the groups – the MOT training group improved on the pre- and post-training MOT task significantly more than did either the passive control or the n-back group [Group x Time interaction, $F(1,31) = 31.4$, $p < .0001$] , while the n-back group improved on the n-back task significantly more than did the MOT training group [Group x Time interaction, $F(1,31) = 45.6$, $p < .0001$]. [Figure 2]

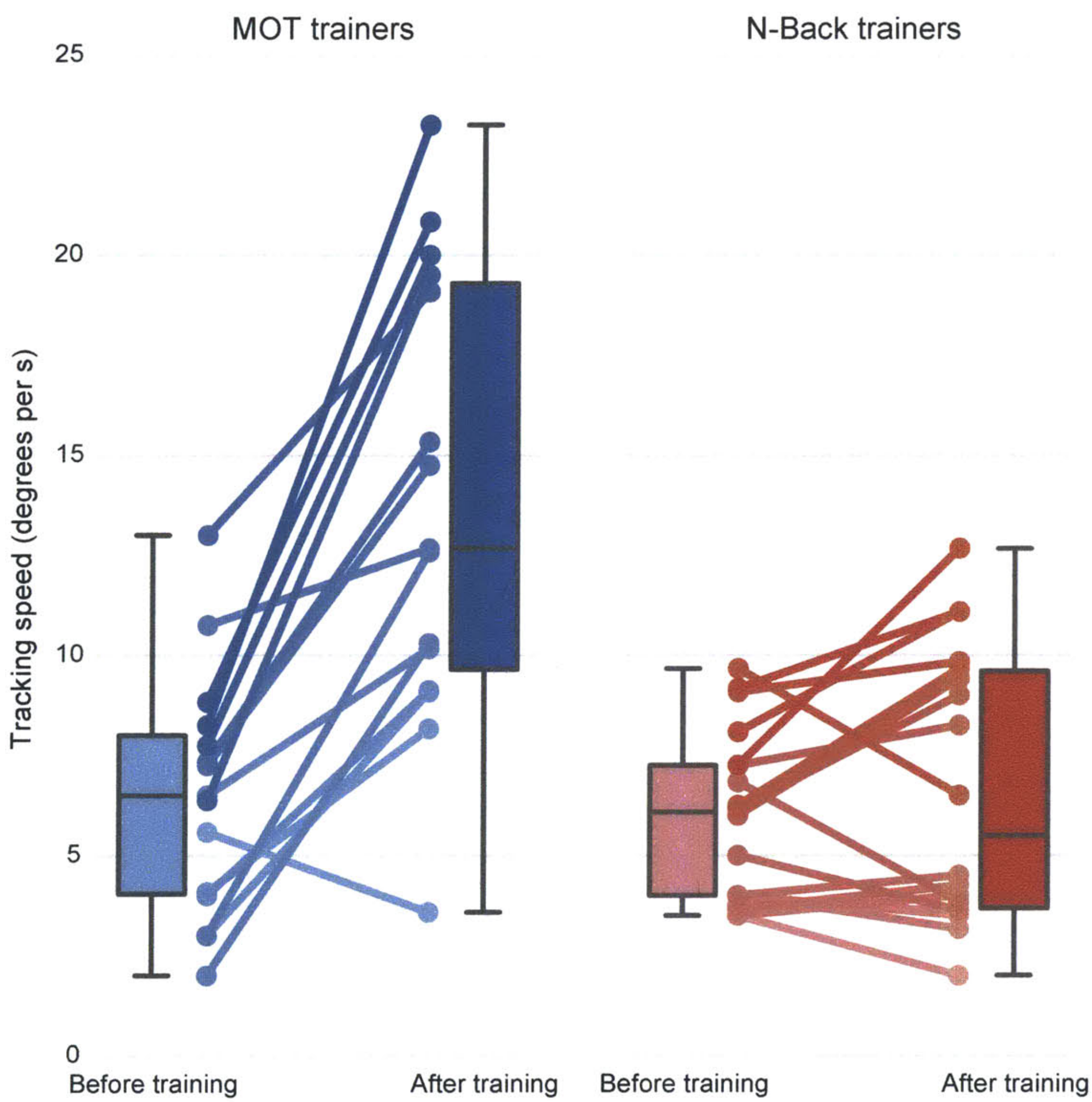


Figure 2: Pre- and post-training tracking speeds for the MOT trainees and the n-back control group. The thresholded speeds at which participants could track four objects are plotted.

This effect continued at the followup measurement, where the trained MOT group continues to outperform the n-back control group [MOT average speed = 14.0, SD = .52; n-back average speed = 8.2, SD = 3.5, $t(14) = 2.3$, $p < .05$].

Adaptive speed training increases the number of tracked items

In the MOT training condition, participants improved from tracking an estimated 2.78 (SD = .45) items during their pre-training number-limit assessment to an estimated 4.16 (SD = .69) items after training, measured at their original tracking speed [$t(14) = 5.91$, $p < .001$]. In contrast, there was no significant change in tracked items measured in the n-back training group [pre-training capacity 2.92, SD .41; post-training capacity 3.14, SD .43; $t(17) = 1.54$, $p > .14$; Group x Time interaction $F(1,31) = 20.2$, $p < .001$].

This effect continued at the followup measurement, where the trained MOT group continues to outperform the n-back control group [MOT average number = 3.7, SD = .80; n-back average number = 3.1, SD = .51, $t(14) = 2.3$, $p < .05$].

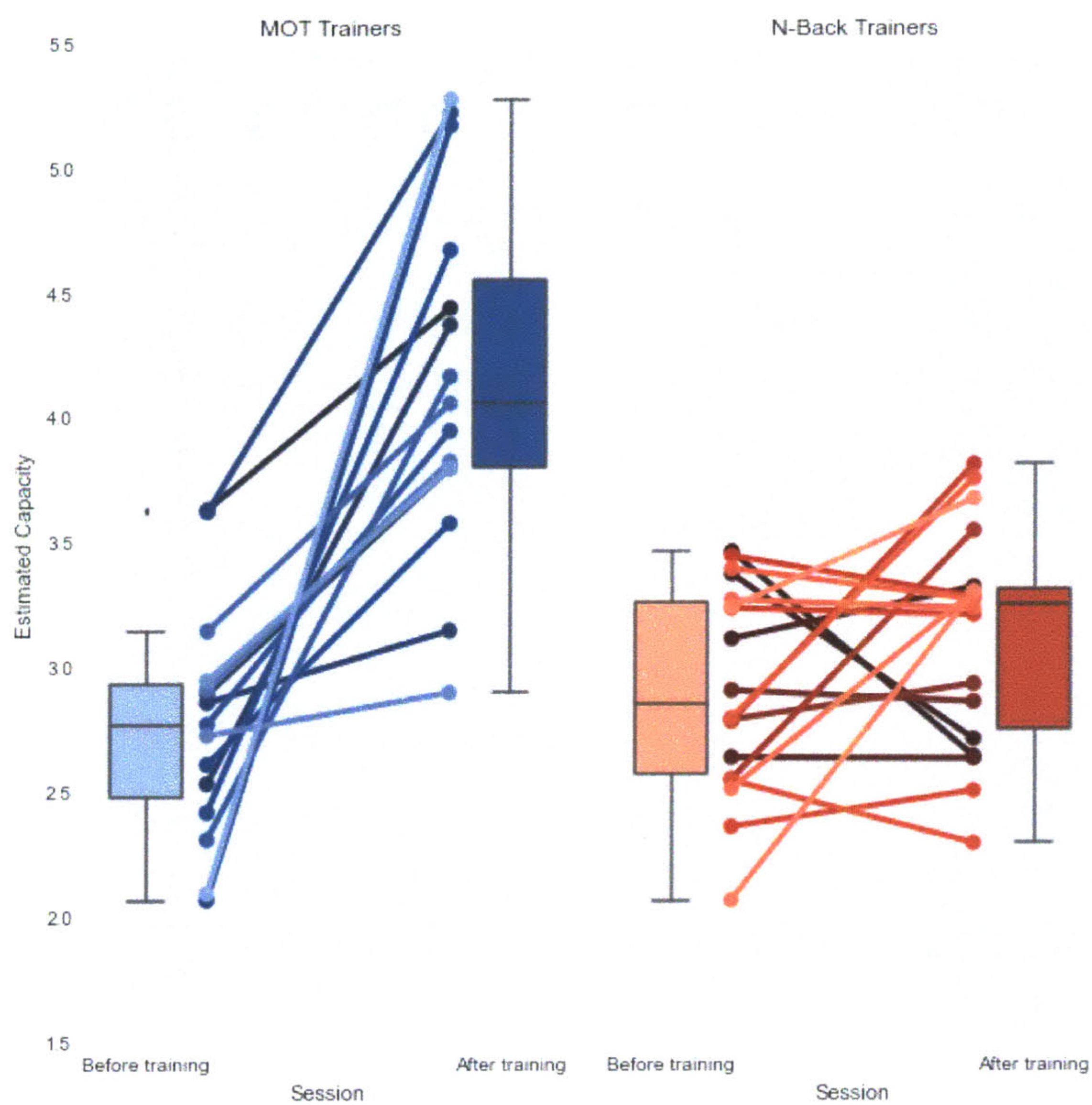


Figure 3a: Pre- and post-training capacity estimates speeds for the MOT trainees and the n-back control group. Error bars represent standard error of the mean.

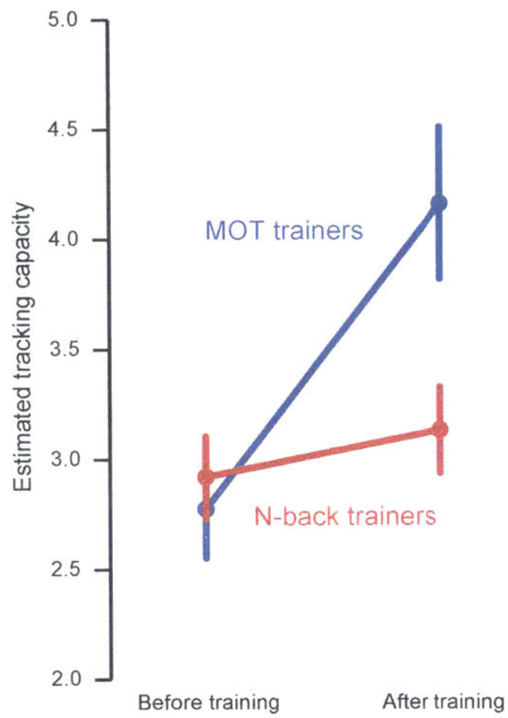


Figure 3b: Pre- and post-training capacity estimates speeds for the MOT trainees and the n-back control group. Error bars represent standard error of the mean.

This improvement in estimated capacity can also be observed in more detail by comparing the performance at each load of the capacity assessment before and after training. [Figures 4 and 5]

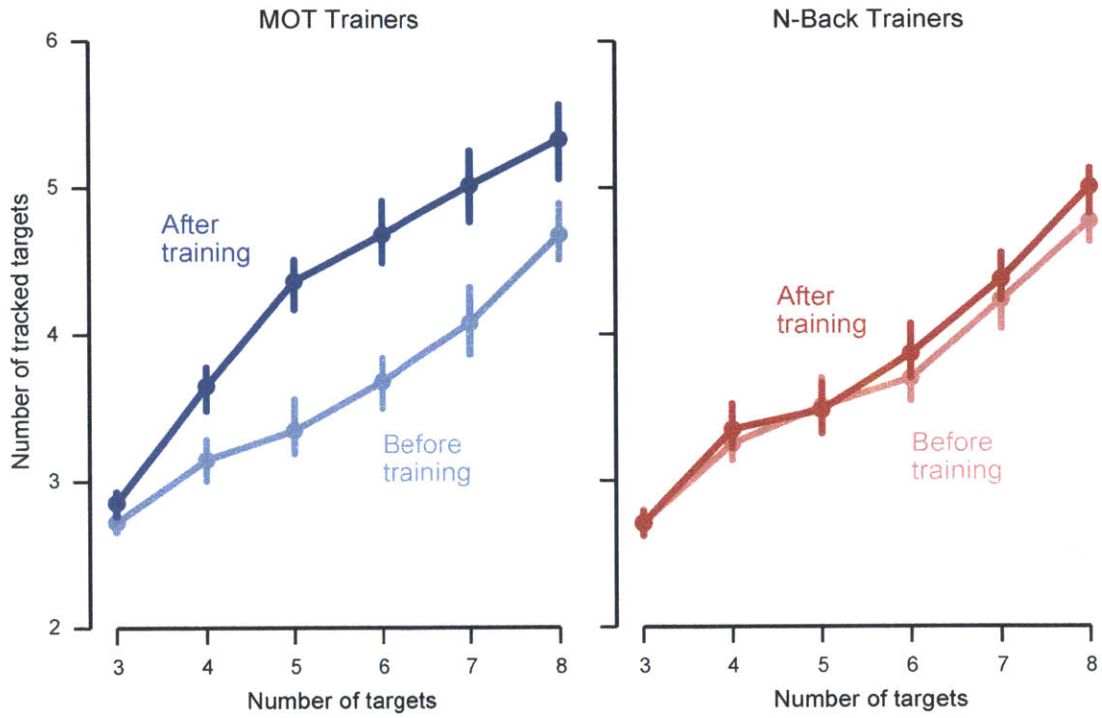


Figure 4: Estimated number of targets tracked, before and after training, for the two trained groups at each MOT load. Error bars represent 95% confidence intervals. Time x group interaction is significant at p < .05 for each target loads between 4 and 8.

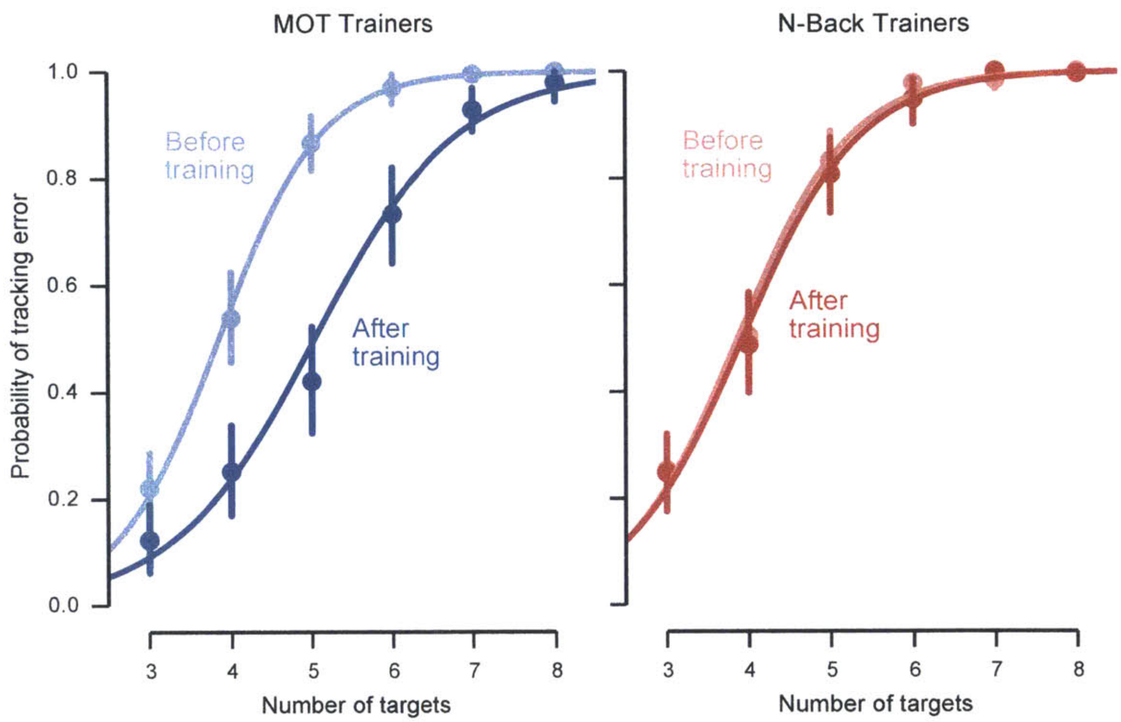


Figure 5: Probability of correctly identifying all targets, before and after training, for both the MOT training group and the dual n-back control group. The curve illustrates the best fitting logistic regression line; error bars indicate 95% confidence intervals for the proportion of error trials at each tracking level.

Capacity gains are linearly related to speed gains

The improvement in capacity is well predicted by the amount of improvement observed during training ($r = .76$, $p < .001$).

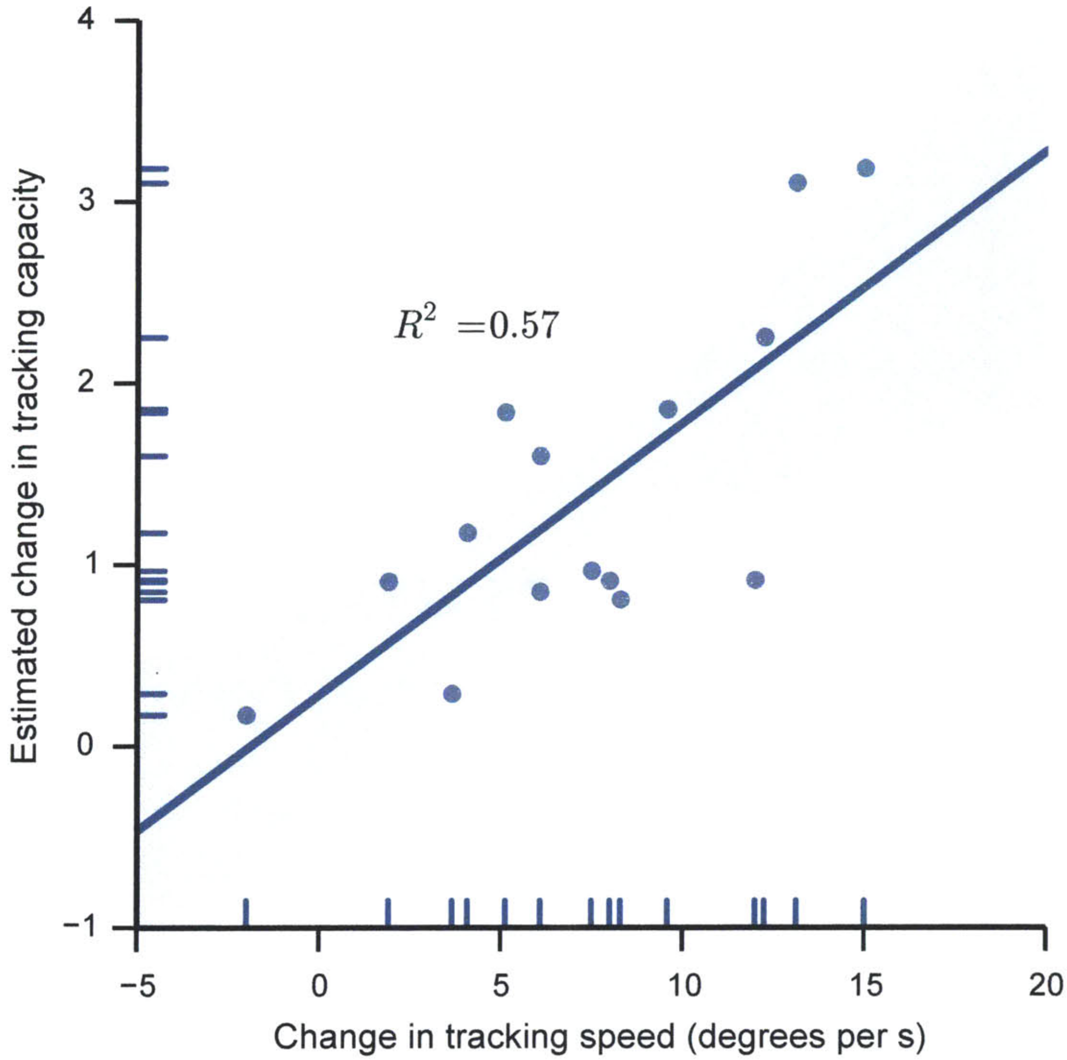


Figure 6: Within the MOT training group, the observed improvement in tracking speed is correlated with the improvement in estimated capacity at the pre-training tracking speed.

Discussion

In this study, we intended to answer two primary questions. First, to what extent is a basic visual attention skill (tracking quickly moving targets) plastic? And secondly, what impact would an improved tracking speed have for the number of items that could subsequently be tracked?

Tracking speed dramatically improves with training

After twenty days of adaptive speed training, participants were able to track four targets significantly faster than they were when they began. The average amount of improvement for each participant over their individual baseline tracking speed was 240%, with continued improvement until at least the last week of training. (To put the magnitude of this improvement in perspective, the average tracking speed for four objects after training was nearly equal to the fastest speed at which participants could track a single object in previous research (Alvarez & Franconeri, 2007)). This experiment may have also underestimated the amount of improvement possible, as some of the faster tracking participants became limited by the physical limits of the laptop display. When objects began moving at more than approximately 22 degrees/second on laptop displays with a 60Hz refresh rate, each object appeared to leave blurred trails behind it as it moved, creating a cluttered and confusing display. It is possible that even greater improvement would be realized with an improved experimental apparatus.

The magnitude of improvement shown here stands in contrast to the one previous research group that has attempted to adaptively train MOT in healthy young adults (using 3 objects instead of 4), who reported observing plateauing performance after 5 training sessions and an average improvement of 40% (Legault et al., 2013). Our data show that multiple object tracking performance can be massively improved with practice, rendering it a more plastic ability than previously assumed. The fact that this ability is so plastic has a number of important implications: for example, it suggests that multiple object tracking performance is not a stable individual characteristic (as assumed, for example, by Drew & Vogel, 2008; Oksama & Hyönä, 2010). Instead, individuals' performance in MOT is likely related not only to their intrinsic characteristics (e.g., individual differences in attentional capacity or in the low-level cortical features that dictate the resolution of attention) but also perhaps to real-world experience with more ecologically relevant counterparts to a tracking task (e.g., juggling?).

In addition, the large performance improvements we see suggest that more research is needed to understand what limits human tracking performance in standard multiple object tracking displays. We found no evidence for an upper bound in tracking performance in our data. Thus, although there are necessarily limits in tracking even in ideal observers (Vul, Frank, Tenenbaum, & Alvarez, 2009), our data suggest that the standard tracking task with 4 objects at moderate speeds may not reveal any immutable capacity limits, since participants are able to continually improve at the task with more training.

How was this improvement achieved?

Improvements after training can be achieved in many ways. Improvements gained solely through the adoption of more efficient strategies may yield impressive results, yet not reflect improvements in any underlying resource pool. (The classic example of this is a trained participant who increased his digit span to an impressive 82 after the development of a set of elaborate encoding strategies, but failed to improve at all on closely related span tasks in which the stimuli were not numeric [Ericcson, Chase, & Faloon, 1980]). It is fair to ask, then, whether the improvements observed in this study could be exclusively explained by changes in task strategy. Although we had no formal tool for evaluating tracking strategies, we did ask each participant at the end of each training day what strategy they were employing and whether it was a change from previous days. Early in the course of training, some shifts were reported, but no strategy changes were reported after approximately the first week of training, although performance continued to improve. Reported strategies varied, with the most commonly reported three strategies being (1) to visualize the tracked dots as corners of a quadrilateral, (2) to attempt to track the center of mass of the four target dots, or (3) to remain fixated on the center fixation cross and track all four target dots in the periphery, without trying to merge the targets into a coherent single object. Based on previous research (Yantis, 1992), shifting to a center-of-mass tracking strategy does yield modest improvements in attentional tracking, but that improvement clearly cannot explain the magnitude of improvements observed here.

People can also track more items.

The second aim of this study was to assess the extent to which training on tracking speed would affect the number of items that could be tracked. This question has important implications for understanding the underlying cognitive architecture that supports working memory performance. Interestingly, we found that training improved not only the speed at which subjects could track four items (the classic limit in discrete models of attentional capacity), but also the number of items they could track. Moreover, across individuals, the increase in numerical capacity was strongly related to the gains in speed.

How does this bear on theories of working memory and visual attention?

The strongest version of a "slot" model would posit that working memory capacity is immutable (at typical speeds), as there is a hard limit imposed by the number of pointers or FINSTs available. On this hypothesis, improving speed should have no effect on the number of items tracked, because it is the number of pointers that functions as the primary bottleneck on performance capacity. The fact that capacity increases with speed training speaks against the simplest version of a fixed, discrete model, converging with extant psychophysical data from (Alvarez & Franconeri, 2007). But does this finding address more sophisticated hybrid models that posit separable limits from a slot-like system underlying capacity limits and resource-like system underlying speed performance? Hybrid models grant that capacity limits are not the sole bottleneck on working memory performance, but posit instead two independent contributors to performance. The present paradigm had the potential to provide strong support for this hypothesis, as it predicts that improving speed performance should have no effect on working memory capacity. If we had failed to observe transfer, it would have been evidence in favor of dissociable, independent systems, suggesting a fixed capacity limiting mechanism.

That we *did* observe transfer speaks against such a fully independent account of speed and capacity limitations. However, our baseline speed did not allow participants to track 4 objects with perfect accuracy. Seeing capacity improvements at such a speed would have provided compelling evidence against a fixed capacity limit of 4 slots, and provided support for a flexible resource model. Because our baseline capacity measurement was relatively low (2.5), post-training capacity was still only slightly higher than 4.25., Thus, it is possible to salvage a hybrid model with a perceptual bottleneck introduced by quickly moving object such that this bottleneck prevents the full utilization of the available slots/pointers. On this account, speed training might remove the bottleneck and enable the full deployment of the fixed resource.

Future research could attempt to rule out this alternative by testing the effects of training when starting capacities are close to 4 items.

One challenge associated with evaluating resource models is that the notion of a resource is often ambiguous. However, a few recent proposals have aimed to bring greater precision to the nature of the underlying resource. For example, (Franconeri, Alvarez, & Cavanagh, 2013) have proposed that the resource is essentially 'cortical real estate'; capacity limits are ultimately determined by competition within map-based representations located in cortex possessing topographic maps of visual space. There is some evidence in support of this view for the multiple object tracking task,(Carlson, Alvarez, & Cavanagh, 2007; Franconeri, Jonathan, & Scimeca, 2010; Shim, Alvarez, & Jiang, 2008) which attributes performance limitations to spatial interference. Franconeri has recently argued that both capacity limits and speed effects might derive from distance-dependent interference effects relating to the spatial receptive fields of occipital and parietal neurons (Franconeri et al., 2013; Franconeri, 2011). When there are many objects, or objects are moving quickly, objects are more likely to collide or crowd one another, meaning that they are more likely to complete within the receptive field of a given neuron. One explanation of training transfer is that speed improvements are supported by increasing the resolution of the "attentional spotlight", perhaps by tightening the receptive fields of neurons involved in maintaining visual representations. If speed and numerical limitations are both driven by crowding, this would explain the results obtained here, specifically the transfer in performance gains across the two metrics.

However, others have argued against a purely spatial account of capacity limits by demonstrating that working memory performance suffers for large numbers of objects even when crowding is held constant and collisions do not occur (Holcombe & Chen, 2012). When objects do not collide and remain widely separated in the visual field, performance still decreases with the number of objects and with the speed (Holcombe et al., 2014). This is evidence against a purely spatial interference account. However, it is consistent with both a pure resource account and a hybrid resource/fixed pointer account.

The current paradigm could be extended to resolve these conflicting findings. Would improvement on speed transfer to increased numerical capacity even in a context where objects do not collide and remain spatially separable? If performance transferred even under these conditions, this would speak against a pure crowding account and provide evidence for a flexible, non-spatial resource.

Conclusion

We found that participants could vastly improve their multiple object tracking performance with training where the speed of objects continually increased. Furthermore, we found this performance improvement transferred to being able to track nearly twice as many objects at a fixed speed. These results stand in contrast to a control group who trained on an adaptive dual n-back training task and showed no significant gains in either tracking speed or attentional tracking capacity. Taken together, these results demonstrate that tracking limits are not fixed, and suggest that speed-limits and number-limits reflect shared underlying constraints on attentional tracking.

Chapter 5: Conclusion

Understanding, and expanding, the limits of human cognition is a key goal of cognitive science. Training studies provide one important way to do this. The trained enhancement of working memory and visual attention has both theoretical implications for understanding the architectures of cognition, as well as practical implications for education and clinical treatment. In particular, transfer of training from one task to another may reveal shared psychological processes or neural systems across domains of cognition. In three experiments presented here, participants underwent a month of intensive training on either a complex working memory task or a visual attention task and results from that training were assessed in the service of three different questions.

Using training studies of working memory and visual attention, I asked **first** to what extent the effects of intensive working memory training can transfer benefits to clinically- and educationally-important external domains; **second**, I explored the neural frameworks supporting human gains within a trained task; and **third**, I examined the plasticity of a visual attention system, and whether training could help dissociate competing hypotheses about the cognitive architecture of visual cognition.

In Experiment 1, I find that humans can make substantial advances in working memory with training but these gains do not transfer beyond the trained tasks. In Experiment 2, I found that neural changes within and across frontoparietal networks supported the observed within-task gains. Specifically, neuroimaging revealed that increased working memory performance was accompanied both by decreases of functional activation in relatively specific portions of the frontoparietal network, and by increases in frontoparietal functional connectivity. Finally, in Experiment 3, I found that object tracking skills within visual attention are surprisingly plastic, and that training on tracking speed improves the number of objects that can be tracked, supporting a model of a shared resource across the two limits in the MOT task.

References

Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2005). Working memory and intelligence: the same or different constructs? *Psychological Bulletin, 131*(1), 30–60. doi:10.1037/0033-2909.131.1.30

Alvarez, G. A., & Franconeri, S. L. (2007). How many objects can you track? Evidence for a resource-limited attentive tracking mechanism. *Journal of Vision, 7*(13), 14.1–10. doi:10.1167/7.13.14

Anguera, J. A., Bernard, J. A., Jaeggi, S. M., Buschkuehl, M., Benson, B. L., Jennett, S., … Seidler, R. D. (2012). The effects of working memory resource depletion and training on sensorimotor adaptation. *Behavioural Brain Research, 228*(1), 107–15. doi:10.1016/j.bbr.2011.11.040

Au, J., Sheehan, E., Tsai, N., Duncan, G. J., Buschkuehl, M., & Jaeggi, S. M. (2014). Improving fluid intelligence with training on working memory: a meta-analysis. *Psychonomic Bulletin & Review*. doi:10.3758/s13423-014-0699-x

Baddeley, A., & Hitch, G. (1974). Working memory. In *The psychology of learning and motivation: Advances in research and theory, Volume 8* (pp. pp. 47–89). New York: Academic Press.

Blackwell, L. S., Trzesniewski, K. H., & Dweck, C. S. (2007). Implicit theories of intelligence predict achievement across an adolescent transition: a longitudinal study and an intervention. *Child Development, 78*(1), 246–63. doi:10.1111/j.1467-8624.2007.00995.x

Blair, C. (2006). How similar are fluid cognition and general intelligence? A developmental neuroscience perspective on fluid cognition as an aspect of human cognitive ability. *Behavioral and Brain Sciences, 29*(02), 109–125.

Boot, W. R., Kramer, A. F., Simons, D. J., Fabiani, M., & Gratton, G. (2008). The effects of video game playing on attention, memory, and executive control. *Acta Psychologica, 129*(3), 387–98. doi:10.1016/j.actpsy.2008.09.005

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10*, 433–436. doi:10.1163/156856897X00357

Braver, T. S., Cohen, J. D., Nystrom, L. E., Jonides, J., Smith, E. E., & Noll, D. C. (1997). A parametric study of prefrontal cortex involvement in human working memory. *NeuroImage, 5*(1), 49–62. doi:10.1006/nimg.1996.0247

Brehmer, Y., Westerberg, H., & Bäckman, L. (2012). Working-memory training in younger and older adults: training gains, transfer, and maintenance. *Frontiers in Human Neuroscience*, *6*(March), 63. doi:10.3389/fnhum.2012.00063

Brown, J. I., Fishco, V. V., & Hanna, G. (1993). *Nelson-Denny reading test*. Riverside Publishing Company.

Burgess, G. C., Gray, J. R., Conway, A. R. A., & Braver, T. S. (2011). Neural mechanisms of interference control underlie the relationship between fluid intelligence and working memory span. *Journal of Experimental Psychology. General*, *140*(4), 674–92. doi:10.1037/a0024695

Buschkuehl, M., Hernandez-Garcia, L., Jaeggi, S. M., Bernard, J. a, & Jonides, J. (2014). Neural effects of short-term training on working memory. *Cognitive, Affective & Behavioral Neuroscience*, *14*, 147–60. doi:10.3758/s13415-013-0244-9

Buschkuehl, M., Jaeggi, S. M., Hutchison, S., Perrig-Chiello, P., Däpp, C., Müller, M., … Perrig, W. J. (2008). Impact of working memory training on memory performance in old-old adults. *Psychology and Aging*, *23*(4), 743–53. doi:10.1037/a0014342

Buschkuehl, M., Jaeggi, S. M., & Jonides, J. (2012). Neuronal effects following working memory training. *Developmental Cognitive Neuroscience*, *2 Suppl 1*, S167–79. doi:10.1016/j.dcn.2011.10.001

Carlson, T. a, Alvarez, G. A., & Cavanagh, P. (2007). Quadrantic deficit reveals anatomical constraints on selection. *Proceedings of the National Academy of Sciences of the United States of America*, *104*, 13496–13500. doi:10.1073/pnas.0702685104

Cattell, R. B. (1936). Is national intelligence declining? *The Eugenics Review*, *28*(3), 181.

Chein, J. M., & Morrison, A. B. (2010). Expanding the mind's workspace: training and transfer effects with a complex working memory span task. *Psychonomic Bulletin & Review*, *17*(2), 193–9. doi:10.3758/PBR.17.2.193

Chooi, W. T., & Thompson, L. A. (2012). Working memory training does not improve intelligence in healthy young adults. *Intelligence*, *40*(6), 531–542.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155–159. doi:10.1037/0033-2909.112.1.155

Cohen, J. D., Perlstein, W. M., Braver, T. S., Nystrom, L. E., Noll, D. C., Jonides, J., & Smith, E. E. (1997). Temporal dynamics of brain activation during a working memory task. *Nature*, *386*(6625), 604–8. doi:10.1038/386604a0

Conway, A. R. ., Cowan, N., Bunting, M. F., Therriault, D. J., & Minkoff, S. R. . (2002). A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence. *Intelligence*, *30*(2), 163–183. doi:10.1016/S0160-2896(01)00096-4

Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, *12*(5), 769–86.

Conway, A. R. A., Kane, M. J., & Engle, R. W. (2003). Working memory capacity and its relation to general intelligence. *Trends in Cognitive Sciences*, *7*(12), 547–552. doi:10.1016/j.tics.2003.10.005

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews. Neuroscience*, *3*(March), 201–215. doi:10.1038/nrn755

Costa, P. T., & McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO PI-R) and Neo Five-Factor Inventory (NEO-FFI)*. Psychological Assessment Resources Odessa, FL.

Culham, J. C., Cavanagh, P., & Kanwisher, N. G. (2001). Attention response functions: Characterizing brain areas using fMRI activation during parametric variations of attentional load. *Neuron*, *32*(4), 737–745. doi:10.1016/S0896-6273(01)00499-8

Dahlin, E., Neely, A. S., Larsson, A., Bäckman, L., Nyberg, L., & Backman, L. (2008). Transfer of learning after updating training mediated by the striatum. *Science (New York, N.Y.)*, *320*(5882), 1510–2. doi:10.1126/science.1155466

Dale, A. M., Fischl, B., & Sereno, M. I. (1999). Cortical surface-based analysis. I. Segmentation and surface reconstruction. *NeuroImage*, *9*(2), 179–94. doi:10.1006/nimg.1998.0395

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, *19*(4), 450–466. doi:10.1016/S0022-5371(80)90312-6

Deary, I., Strand, S., Smith, P., & Fernandes, C. (2007). Intelligence and educational achievement. *Intelligence*, *35*(1), 13–21. doi:10.1016/j.intell.2006.02.001

Deary, I., Whalley, L., Lemmon, H., Crawford, J. ., & Starr, J. M. (2000). The Stability of Individual Differences in Mental Ability from Childhood to Old Age: Follow-up of the 1932 Scottish Mental Survey. *Intelligence*, *28*(1), 49–55. doi:10.1016/S0160-2896(99)00031-8

Drew, T., & Vogel, E. K. (2008). Neural measures of individual differences in selecting and tracking multiple moving objects. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *28*(16), 4183–91. doi:10.1523/JNEUROSCI.0556-08.2008

Duckworth, A. L., Peterson, C., Matthews, M. D., & Kelly, D. R. (2007). Grit: perseverance and passion for long-term goals. *Journal of Personality and Social Psychology*, *92*(6), 1087–101. doi:10.1037/0022-3514.92.6.1087

Duckworth, A. L., & Quinn, P. D. (2009). Development and validation of the short grit scale (grit-s). *Journal of Personality Assessment*, *91*(2), 166–74. doi:10.1080/00223890802634290

Duncan, J. (2010). The multiple-demand (MD) system of the primate brain: mental programs for intelligent behaviour. *Trends in Cognitive Sciences*. doi:10.1016/j.tics.2010.01.004

Dweck, C. (2006). *Mindset: The New Psychology of Success* (p. 288). Random House.

Dweck, C. S. (2000). *Self-theories: Their role in motivation, personality, and development.* Psychology Press.

Elbert, T., Pantev, C., Wienbruch, C., Rockstroh, B., & Taub, E. (1995). Increased Cortical Representation of the Fingers of the Left Hand in String Players. *Science*, *270*(5234), 305–307. doi:10.1126/science.270.5234.305

Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science*, *11*(1), 19–23.

Engle, R. W., Kane, M. J., & Tuholski, S. W. (1999). Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence, and functions of the prefrontal cortex. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 102–134). Cambridge University Press.

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. A. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology. General*, *128*(3), 309–331. doi:10.1037/0096-3445.128.3.309

Ericcson, K. A., Chase, W. G., & Faloon, S. (1980). Acquisition of a memory skill. *Science (New York, N.Y.)*, *208*, 1181–1182. doi:10.1126/science.7375930

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. doi:10.3758/BF03193146

Finn, A. S., Kraft, M. a, West, M. R., Leonard, J. a, Bish, C. E., Martin, R. E., … Gabrieli, J. D. E. (2014). Cognitive skills, student achievement tests, and schools. *Psychological Science*, *25*(3), 736–44. doi:10.1177/0956797613516008

Franconeri, S. L. (2011). The nature and status of visual resources. *Oxford Handbook of Cognitive Psychology*, 1–16. doi:10.1093/oxfordhb/9780195376746.013.0010

Franconeri, S. L., Alvarez, G. A., & Cavanagh, P. (2013). Flexible cognitive resources: Competitive content maps for attention and memory. *Trends in Cognitive Sciences*, *17*(3), 134–141. doi:10.1016/j.tics.2013.01.010

Franconeri, S. L., Jonathan, S. V, & Scimeca, J. M. (2010). Tracking multiple objects is limited only by object spacing, not by speed, time, or capacity. *Psychological Science : A Journal of the American Psychological Society / APS*, *21*, 920–925. doi:10.1177/0956797610373935

Funahashi, S., Bruce, C. J., & Goldman-Rakic, P. S. (1989). Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. *Journal of Neurophysiology*, *61*(2), 331–49.

Garavan, H., Kelley, D., Rosen, A., Rao, S. M., & Stein, E. A. (2000). Practice-related functional activation changes in a working memory task. *Microscopy Research and Technique*, *51*(January), 54–63. doi:10.1002/1097-0029(20001001)51:1<54::AID-JEMT6>3.0.CO;2-J

Gaspar, J. G., Neider, M. B., Simons, D. J., McCarley, J. S., & Kramer, A. F. (2013). Change Detection: Training and Transfer. *PLoS ONE*, *8*(6), 1–7. doi:10.1371/journal.pone.0067781

Gathercole, S. E., Pickering, S. J., Knight, C., & Stegmann, Z. (2004). Working memory skills and educational attainment: evidence from national curriculum assessments at 7 and 14 years of age. *Applied Cognitive Psychology*, *18*(1), 1–16. doi:10.1002/acp.934

Gorgolewski, K., Burns, C. D., Madison, C., Clark, D., Halchenko, Y. O., Waskom, M. L., & Ghosh, S. S. (2011). Nipype: a flexible, lightweight and extensible neuroimaging data processing framework in python. *Frontiers in Neuroinformatics*, *5*(August), 13. doi:10.3389/fninf.2011.00013

Gottfredson, L. S. (1997, February). Why g matters: The complexity of everyday life. *Intelligence*. doi:10.1016/S0160-2896(97)90014-3

Gray, J. R., Chabris, C. F., & Braver, T. S. (2003). Neural mechanisms of general fluid intelligence. *Nature Neuroscience*, *6*(3), 316–22. doi:10.1038/nn1014

Green, C. S., & Bavelier, D. (2006). Enumeration versus multiple object tracking: the case of action video game players. *Cognition*, *101*(1), 217–45. doi:10.1016/j.cognition.2005.10.004

Greve, D. N., & Fischl, B. (2009). Accurate and robust brain image alignment using boundary-based registration. *NeuroImage*, *48*(1), 63–72. doi:10.1016/j.neuroimage.2009.06.060

Haier, R. J., Siegel, B., Tang, C., Abel, L., & Buchsbaum, M. S. (1992). Intelligence and changes in regional cerebral glucose metabolic rate following learning. *Intelligence*, *16*(3-4), 415–426. doi:10.1016/0160-2896(92)90018-M

Haier, R. J., Siegel, B. V, MacLachlan, A., Soderling, E., Lottenberg, S., & Buchsbaum, M. S. (1992). Regional glucose metabolic changes after learning a complex visuospatial/motor task: a positron emission tomographic study. *Brain Research*, *570*(1-2), 134–43.

Heitz, R. P., Redick, T. S., & Hambrick, D. Z. (2006). Working memory , executive function , and general fluid intelligence are not the same Clarifying process versus structure in human intelligence : Stop talking about fluid and crystallized, 135–136.

Hempel, A., Giesel, F. L., Garcia Caraballo, N. M., Amann, M., Meyer, H., Wüstenberg, T., … Schröder, J. (2004). Plasticity of cortical activation related to working memory during training. *The American Journal of Psychiatry*, *161*(April), 745–747. doi:10.1176/appi.ajp.161.4.745

Holcombe, A. O., & Chen, W.-Y. (2012). Exhausting attentional tracking resources with a single fast-moving object. *Cognition*, *123*(2), 218–28. doi:10.1016/j.cognition.2011.10.003

Holcombe, A. O., Chen, W.-Y., & Howe, P. D. L. (2014). Object tracking: Absence of long-range spatial interference supports resource theories. *Journal of Vision*, *14*, 1–39. doi:10.1167/14.6.1

Holmes, J., Gathercole, S. E., & Dunning, D. L. (2009). Adaptive training leads to sustained enhancement of poor working memory in children. *Developmental Science*, *12*(4), F9–15. doi:10.1111/j.1467-7687.2009.00848.x

Hulleman, J. (2005). The mathematics of multiple object tracking: From proportions correct to number of objects tracked. *Vision Research*, *45*, 2298–2309. doi:10.1016/j.visres.2005.02.016

Jaeggi, S. M., Buschkuehl, M., Etienne, A., Ozdoba, C., Perrig, W. J., & Nirkko, A. C. (2007). On how high performers keep cool brains in situations of cognitive overload. *Cognitive, Affective & Behavioral Neuroscience*, *7*(2), 75–89.

Jaeggi, S. M., Buschkuehl, M., Jonides, J., & Perrig, W. J. (2008). Improving fluid intelligence with training on working memory. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(19), 6829–33. doi:10.1073/pnas.0801268105

Jaeggi, S. M., Buschkuehl, M., Jonides, J., & Shah, P. (2011). Short- and long-term benefits of cognitive training. *Proceedings of the National Academy of Sciences*, *108*(25), 10081–6. doi:10.1073/pnas.1103228108

Jaeggi, S. M., Studer-Luethi, B., Buschkuehl, M., Su, Y.-F., Jonides, J., & Perrig, W. J. (2010). The relationship between n-back performance and matrix reasoning — implications for training and transfer. *Intelligence*, *38*(6), 625–635. doi:10.1016/j.intell.2010.09.001

Jansma, J. M., Ramsey, N. F., Slagter, H. A., & Kahn, R. S. (2001). Functional anatomical correlates of controlled and automatic processing. *Journal of Cognitive Neuroscience*, *13*(6), 730–743.

Jaušovec, N., & Jaušovec, K. (2012). Working memory training: improving intelligence--changing brain activity. *Brain and Cognition*, *79*(2), 96–106. doi:10.1016/j.bandc.2012.02.007

Jenkinson, M., Beckmann, C. F., Behrens, T. E. J., Woolrich, M. W., & Smith, S. M. (2012). FSL. *NeuroImage*, *62*(2), 782–90. doi:10.1016/j.neuroimage.2011.09.015

Jolles, D. D., Grol, M. J., Van Buchem, M. A., Rombouts, S. A. R. B., & Crone, E. A. (2010). Practice effects in the brain: Changes in cerebral activation after working memory practice depend on task demands. *NeuroImage*, *52*(2), 658–668. doi:10.1016/j.neuroimage.2010.04.028

Jones, E., Oliphant, T., & Peterson, P. (2001). SciPy: Open source scientific tools for Python.

Jonides, J. (2004). How does practice makes perfect? *Nature Neuroscience*, *7*(1), 10–11.

Judge, T. A., Hurst, C., & Simon, L. S. (2009). Does it pay to be smart, attractive, or confident (or all three)? Relationships among general mental ability, physical attractiveness, core self-evaluations, and income. *The Journal of Applied Psychology*, *94*(3), 742–755. doi:10.1037/a0015497

Jung, R. E., & Haier, R. J. (2007). The Parieto-Frontal Integration Theory (P-FIT) of intelligence: converging neuroimaging evidence. *The Behavioral and Brain Sciences*, *30*, 135–154; discussion 154–187. doi:10.1017/S0140525X07001185
Kane, M. J., & Engle, R. W. (2002). The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: an individual-differences perspective. *Psychonomic Bulletin & Review*, *9*(4), 637–71.
Kane, M. J., Hambrick, D. Z., & Conway, A. R. A. (2005). Working memory capacity and fluid intelligence are strongly related constructs: comment on Ackerman, Beier, and Boyle (2005). *Psychological Bulletin*, *131*(1), 66–71; author reply 72–5. doi:10.1037/0033-2909.131.1.66
Kelly, C., Foxe, J. J., & Garavan, H. (2006). Patterns of normal human brain plasticity after practice and their implications for neurorehabilitation. *Archives of Physical Medicine and Rehabilitation*, *87*(12 Suppl 2), S20–9. doi:10.1016/j.apmr.2006.08.333
Kelly, C., & Garavan, H. (2005). Human functional neuroimaging of brain changes associated with practice. *Cerebral Cortex (New York, N.Y. : 1991)*, *15*(8), 1089–102. doi:10.1093/cercor/bhi005
Klingberg, T. (2010). Training and plasticity of working memory. *Trends in Cognitive Sciences*, *14*(7), 317–24. doi:10.1016/j.tics.2010.05.002
Klingberg, T., Fernell, E., Olesen, P. J., Johnson, M., Gustafsson, P., Dahlström, K., … Westerberg, H. (2005). *Computerized training of working memory in children with ADHD--a randomized, controlled trial. Journal of the American Academy of Child and Adolescent Psychiatry* (Vol. 44, pp. 177–186). doi:10.1097/00004583-200502000-00010
Klingberg, T., Forssberg, H., & Westerberg, H. (2002). *Training of working memory in children with ADHD. Journal of clinical and experimental neuropsychology* (Vol. 24, pp. 781–791). doi:10.1076/jcen.24.6.781.8395
Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working-memory capacity?!*. *Intelligence*, *14*(4), 389–433.
Landau, S. M., Garavan, H., Schumacher, E. H., & D'Esposito, M. (2007). Regional specificity and practice: dynamic changes in object and spatial working memory. *Brain Research*, *1180*, 78–89. doi:10.1016/j.brainres.2007.08.057
Landau, S. M., Schumacher, E. H., Garavan, H., Druzgal, T. J., & D'Esposito, M. (2004). A functional MRI study of the influence of practice on component processes of working memory. *NeuroImage*, *22*(1), 211–21. doi:10.1016/j.neuroimage.2004.01.003
Legault, I., Allard, R., & Faubert, J. (2013). Healthy older observers show equivalent perceptual-cognitive training benefits to young adults for multiple object tracking. *Frontiers in Psychology*, *4*(June), 323. doi:10.3389/fpsyg.2013.00323
Loosli, S. V., Buschkuehl, M., Perrig, W. J., & Jaeggi, S. M. (2012). Working memory training improves reading processes in typically developing children. *Child Neuropsychology : A Journal on Normal and Abnormal Development in Childhood and Adolescence*, *18*(1), 62–78. doi:10.1080/09297049.2011.575772
Lu, Z. L., Chu, W., Dosher, B. A., & Lee, S. (2005). Perceptual learning of Gabor orientation identification in visual periphery: Complete inter-ocular transfer of learning mechanisms. *Vision Research*, *45*, 2500–2510. doi:10.1016/j.visres.2005.04.009
McCabe, D. P., Roediger III, H. L., McDaniel, M. A., Balota, D. A., & Hambrick, D. Z. (2010). The relationship between working memory capacity and executive functioning: Evidence for a common executive attention construct. *Neuropsychology*, *24*(2), 222.

Mcnab, F., & Klingberg, T. (2008). Prefrontal cortex and basal ganglia control access to working memory. *Nature Neuroscience*, *11*(1), 103–7. doi:10.1038/nn2024

McNab, F., Varrone, A., Farde, L., Jucaite, A., Bystritsky, P., Forssberg, H., & Klingberg, T. (2009). Changes in cortical dopamine D1 receptor binding associated with cognitive training. *Science (New York, N.Y.)*, *323*(5915), 800–2. doi:10.1126/science.1166102

Melby-Lervåg, M., & Hulme, C. (2013). Is working memory training effective? A meta-analytic review. *Developmental Psychology*, *49*(2), 270–91. doi:10.1037/a0028228

Miller, E. K., Erickson, C. a, & Desimone, R. (1996). Neural mechanisms of visual working memory in prefrontal cortex of the macaque. *Journal of Neuroscience*, *16*(16), 5154–5167.

Miller, G. A. (1956). The magical number seven plus or minus two: some limits on our capacity for processing information. *Psychological Review*, *63*(2), 81–97.

Mischel, W., Shoda, Y., & Rodriguez, M. (1989). Delay of gratification in children. *Science*, *244*(4907), 933 –938. doi:10.1126/science.2658056

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex Frontal Lobe Tasks: A Latent Variable Analysis. *Cognitive Psychology*, *41*(1), 49–100.

Moody, D. E. (2009, July). Can intelligence be increased by training on a task of working memory? *Intelligence*, *37*(4), 327–328. doi:10.1016/j.intell.2009.04.005

Neisser, U., Boodoo, G., Bouchard, Thomas J., J., Boykin, a. W., Brody, N., Ceci, S. J., … Urbina, S. (1996). Intelligence: Knowns and unknowns. *The American Psychologist*, *51*(2), 77–101. doi:10.1037/0003-066X.51.2.77

Nisbett, R. E., Aronson, J., Blair, C., Dickens, W., Flynn, J., Halpern, D. F., & Turkheimer, E. (2012). Intelligence: new findings and theoretical developments. *The American Psychologist*, *67*(2), 130–59. doi:10.1037/a0026699

Novick, J. M. J., Hussey, E., Teubner-Rhodes, S., Harbison, J. I., & Bunting, M. F. (2013). Clearing the garden-path: Improving sentence processing through cognitive control training. *Language and Cognitive Processes*, *0*(April 2013), 1–44. doi:10.1080/01690965.2012.758297

Oksama, L., & Hyönä, J. (2010). Is multiple object tracking carried out automatically by an early vision mechanism independent of higher-order cognition? An individual difference approach.

Olesen, P. J., Westerberg, H., & Klingberg, T. (2004). Increased prefrontal and parietal activity after training of working memory. *Nature Neuroscience*, *7*(1), 75–9. doi:10.1038/nn1165

Ostensson, M.-L., Hirvikoski, T., Westerberg, H., Klingberg, T., Jacobaeus, H., Clevberger, P., & Bartfai, a. (2007). Computerized working memory training after stroke--a pilot study. *Brain Injury : [BI]*, *21*(1), 21–9. doi:10.1080/02699050601148726

Owen, A. M., Hampshire, A., Grahn, J. A., Stenton, R., Dajani, S., Burns, A. S., … Ballard, C. G. (2010). Putting brain training to the test. *Nature*, *465*(7299), 775–8. doi:10.1038/nature09042

Owen, A. M., McMillan, K. M., Laird, A. R., & Bullmore, E. (2005). N-back working memory paradigm: A meta-analysis of normative functional neuroimaging studies. *Human Brain Mapping*, *25*(December 2004), 46–59. doi:10.1002/hbm.20131

Pascual-Leone, A., Amedi, A., Fregni, F., & Merabet, L. B. (2005). The plastic human brain cortex. *Annual Review of Neuroscience*, *28*, 377–401. doi:10.1146/annurev.neuro.27.070203.144216

Petersen, S. E., van Mier, H., Fiez, J. A., & Raichle, M. E. (1998). The effects of practice on the functional anatomy of task performance. *Proc Natl Acad Sci USA*, *95*(February), 853–860.

Poldrack, R. A. (2000). Imaging brain plasticity: conceptual and methodological issues--a theoretical review. *NeuroImage*, *12*, 1–13. doi:10.1006/nimg.2000.0596

Poldrack, R. A. (2014). Is "efficiency" a useful concept in cognitive neuroscience? *Developmental Cognitive Neuroscience*, 1–6. doi:10.1016/j.dcn.2014.06.001

Poldrack, R. A., Desmond, J. E., Glover, G. H., & Gabrieli, J. D. E. (1998). The neural basis of visual skill learning: An fMRI study of mirror reading. *Cerebral Cortex*, *8*(1976), 1–10. doi:10.1093/cercor/8.1.1

Pylyshyn, Z. W., & Storm, R. W. (1988). Tracking multiple independent targets: Evidence for a parallel tracking mechanism*. *Spatial Vision*, *3*(3), 179–197. doi:10.1163/156856888X00122

Ramsden, S., Richardson, F. M., Josse, G., Thomas, M. S. C., Ellis, C., Shakeshaft, C., … Price, C. J. (2011). Verbal and non-verbal intelligence changes in the teenage brain. *Nature*, *479*(7371), 113–116. doi:10.1038/nature10514

Raven, JC, Court, JH, & Raven, J. (1998). *Raven's Progressive Matrices*. Oxford: Oxford Psychological Press.

Redick, T. S., Shipstead, Z., Harrison, T. L., Hicks, K. L., Fried, D. E., Hambrick, D. Z., … Engle, R. W. (2012). No Evidence of Intelligence Improvement After Working Memory Training: A Randomized, Placebo-Controlled Study. *Journal of Experimental Psychology. General*, 1–86. doi:10.1037/a0029082

Sakai, K., Hikosaka, O., Miyauchi, S., Takino, R., Sasaki, Y., & Pütz, B. (1998). Transition of brain activation from frontal to parietal areas in visuomotor sequence learning. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *18*(5), 1827–40.

Sayala, S., Sala, J. B., & Courtney, S. M. (2006). Increased neural efficiency with repeated performance of a working memory task is information-type dependent. *Cerebral Cortex*, *16*(May), 609–617. doi:10.1093/cercor/bhj007

Schneiders, J., Opitz, B., Krick, C., & Mecklinger, A. (2011). Separating intra-modal and across-modal training effects in visual working memory: an fMRI investigation. *Cerebral Cortex (New York, N.Y. : 1991)*, *21*(11), 2555–64. doi:10.1093/cercor/bhr037

Schneiders, J., Opitz, B., Tang, H., Deng, Y., Xie, C., Li, H., & Mecklinger, A. (2012). The impact of auditory working memory training on the fronto-parietal working memory network. *Frontiers in Human Neuroscience*, *6*(June), 173. doi:10.3389/fnhum.2012.00173

Scholl, B. (2009). *Computation, Cognition, and Pylyshyn* (p. 344). MIT Press.

Schunk, D. H. (2007). *Learning Theories: An Educational Perspective* (5th ed., p. 592). Prentice Hall.

Schwab, S., & Memmert, D. (2012). The impact of a sports vision training program in youth field hockey players. *Journal of Sports Science & Medicine*, *11*(4), 624–31.

Schweizer, S., Grahn, J., Hampshire, A., Mobbs, D., & Dalgleish, T. (2013). Training the emotional brain: improving affective control through emotional working memory training. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *33*(12), 5301–11. doi:10.1523/JNEUROSCI.2593-12.2013

Shadmehr, R., & Holcomb, H. H. (1997). Neural correlates of motor memory consolidation. *Science (New York, N.Y.)*, *277*(5327), 821–5. doi:10.1126/science.277.5327.821

Shim, W. M., Alvarez, G. A., & Jiang, Y. V. (2008). Spatial separation between targets constrains maintenance of attention on multiple objects. *Psychonomic Bulletin & Review, 15*(2), 390–397. doi:10.3758/PBR.15.2.390

Shim, W. M., Alvarez, G. A., Vickery, T. J., & Jiang, Y. V. (2010). The number of attentional foci and their precision are dissociated in the posterior parietal cortex. *Cerebral Cortex (New York, N.Y. : 1991), 20*(6), 1341–9. doi:10.1093/cercor/bhp197

Shipstead, Z., Redick, T. S., & Engle, R. W. (2012). Is working memory training effective? *Psychological Bulletin, 138*(4), 628–54. doi:10.1037/a0027473

Silver, M. a, & Kastner, S. (2009). Topographic maps in human frontal and parietal cortex. *Trends in Cognitive Sciences*. doi:10.1016/j.tics.2009.08.005

Smith, E. E., & Jonides, J. (1997, June). Working memory: a view from neuroimaging. *Cognitive Psychology*, pp. 5–42. doi:10.1006/cogp.1997.0658

Smith, S., & Brady, J. (1997). SUSAN—a new approach to low level image processing. *International Journal of Computer Vision, 23*(1), 45–78. doi:10.1023/A:1007963824710

Smith, S. M., Jenkinson, M., Woolrich, M. W., Beckmann, C. F., Behrens, T. E. J., Johansen-Berg, H., ... Matthews, P. M. (2004). Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage, 23 Suppl 1*, S208–19. doi:10.1016/j.neuroimage.2004.07.051

Sreenivasan, K. K., Curtis, C. E., & D'Esposito, M. (2014). Revisiting the role of persistent neural activity during working memory. *Trends in Cognitive Sciences, 18*, 82–89. doi:10.1016/j.tics.2013.12.001

Sternberg, R. J. (2008). Increasing fluid intelligence is possible after all. *Proceedings of the National Academy of Sciences, 105*(19), 6791. doi:10.1073/pnas.0803396105

Studer-Luethi, B., Jaeggi, S. M., Buschkuehl, M., & Perrig, W. J. (2012). Influence of neuroticism and conscientiousness on working memory training outcome. *Personality and Individual Differences, 53*(1), 44–49. doi:10.1016/j.paid.2012.02.012

Takeuchi, H., Sekiguchi, A., Taki, Y., Yokoyama, S., Yomogida, Y., Komuro, N., ... Kawashima, R. (2010). Training of working memory impacts structural connectivity. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience, 30*(9), 3297–303. doi:10.1523/JNEUROSCI.4611-09.2010

Takeuchi, H., Taki, Y., Hashizume, H., Sassa, Y., Nagase, T., Nouchi, R., & Kawashima, R. (2011). Effects of training of processing speed on neural systems. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience, 31*(34), 12139–48. doi:10.1523/JNEUROSCI.2948-11.2011

Thompson, T. W., Waskom, M. L., Garel, K.-L. a, Cardenas-Iniguez, C., Reynolds, G. O., Winter, R., ... Gabrieli, J. D. E. (2013). Failure of working memory training to enhance cognition or intelligence. *PloS One, 8*(5), e63614. doi:10.1371/journal.pone.0063614

Ungerleider, L. G., Doyon, J., & Karni, A. (2002). Imaging brain plasticity during motor skill learning. *Neurobiology of Learning and Memory, 78*(3), 553–64. doi:10.1006/nlme.2002.4091

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods, 37*(3), 498–505. doi:10.3758/BF03192720

Vul, E., Frank, M. C., Tenenbaum, J. B., & Alvarez, G. A. (2009). Explaining human multiple object tracking as resource-constrained approximate inference in a dynamic probabilistic model. *Advances in Neural Information Processing Systems*.

Wager, T. D., & Smith, E. E. (2003). Neuroimaging studies of working memory: a meta-analysis. *Cognitive, Affective & Behavioral Neuroscience*, *3*(4), 255–74.
Waskom, M. L., Kumaran, D., Gordon, A. M., Rissman, J., & Wagner, A. D. (2014). Frontoparietal Representations of Task Context Support the Flexible Control of Goal-Directed Cognition. *Journal of Neuroscience*, *34*, 10743–10755. doi:10.1523/JNEUROSCI.5282-13.2014
Wechsler, D. (1999). *Wechsler abbreviated scale of intelligence*. Psychological Corporation.
Wechsler, D. (2005). Wechsler Adult Intelligence Scale-III (WAIS-III).
Whalley, L., & Deary, I. (2001). Longitudinal cohort study of childhood IQ and survival up to age 76. *BMJ*, *322*(7290), 819.
Wickens, T. D. (2001). *Elementary Signal Detection Theory. Elementary Signal Detection Theory* (1st ed., Vol. 1, p. 288). Oxford University Press, USA.
Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). Woodcock-Johnson III Battery. *Itasca, IL: Riverside Publishing Company*.
Woolrich, M. W., Ripley, B. D., Brady, M., & Smith, S. M. (2001). Temporal autocorrelation in univariate linear modeling of FMRI data. *NeuroImage*, *14*(6), 1370–86. doi:10.1006/nimg.2001.0931
Yantis, S. (1992). Multielement visual tracking: attention and perceptual organization. *Cognitive Psychology*, *24*, 295–340. doi:10.1016/0010-0285(92)90010-Y
Yeo, B. T. T., Krienen, F. M., Sepulcre, J., Sabuncu, M. R., Lashkari, D., Hollinshead, M., … Buckner, R. L. (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of Neurophysiology*, *106*(3), 1125–65. doi:10.1152/jn.00338.2011
Yoo, S.-S., Paralkar, G., & Panych, L. P. (2004). Neural substrates associated with the concurrent performance of dual working memory tasks. *The International Journal of Neuroscience*, *114*(August 2003), 613–631. doi:10.1080/00207450490430561