

**Moderators of Cognitive Training:  
Individual Differences and Motivational Factors**

by

Benjamin Katz

A dissertation submitted in partial fulfillment  
of the requirements for the degree of  
Doctor of Philosophy  
(Education & Psychology)  
in the University of Michigan  
2017

Doctoral Committee:

Professor Priti R. Shah, Chair  
Professor Pamela Davis-Kean  
Professor John Jonides  
Professor Fredrick J. Morrison

Benjamin Katz

benkatz@umich.edu

ORCID iD: 0000-0002-5612-7540

© Benjamin Katz 2017

## ACKNOWLEDGEMENTS

I could not have done this without Priti Shah, my primary advisor, whose mentorship and support have been invaluable, as well as John Jonides, my secondary advisor, who has always reminded me to take “little steps for little feet.” I am deeply indebted to Pam Davis-Kean and Fred Morrison, my other two committee members, who have guided my research and thinking through countless conversations and lab meetings. I feel lucky to have collaborated with a variety of other faculty members here and outside Michigan who have helped to mold my perspective on cognitive training research, including Patti Reuter-Lorenz, Alison Miller, Susanne Jaeggi, and Martin Buschkuhl. I can say the same of countless graduate students, including Jacky Au and Masha Jones at Irvine and Rebecca Rhodes and Tessa Abagis at Michigan. I am thankful for the help and support from my friends and colleagues within the Basic and Applied Cognition Laboratory and the Cognitive Neuroimaging Laboratory, including Jessica Glazier, Chelsea Zabel, Amira Ibrahim, and many others. My thinking on these topics has also been heavily influenced by many conversations with Dave Meyer, who I can always count on to have a witty Youtube rejoinder to my email messages. The many students and faculty within the LIFE Academy have provided invaluable feedback on my work, and I would not have made it through the last six months of my time here without the help of my fellow Murder players, and Tessa Abagis, Myrna Cintron-Valentin, Tiffany Jantz, Jordan Meyer, Mike Kling, and Klarg. I would not have even become a graduate student if not for my colleagues at Enspire Learning in Austin, Texas and at Lumos Labs in San Francisco, California, and I wouldn’t have stuck with it if not

for the support from my wonderful family, and, of course, Pilar. Finally, I owe thanks my grandmothers Florence and Lee, who were the first to help me understand that this work was worth pursuing, two decades before I began my PhD.

## TABLE OF CONTENTS

<b>ACKNOWLEDGEMENTS.....</b>	<b>ii</b>
<b>LIST OF TABLES.....</b>	<b>v</b>
<b>LIST OF FIGURES.....</b>	<b>vi</b>
<b>ABSTRACT.....</b>	<b>vii</b>
<b>CHAPTER 1: The Origins of the “American Question”.....</b>	<b>1</b>
Introduction.....	1
What is Cognitive Training?.....	2
A Problem of Semantics?.....	4
Historical Interventions in the Public Eye.....	6
The Example of Catherine Aiken.....	9
The Present Issue.....	17
<b>CHAPTER 2: A Map of the Problem: Individual Differences and Motivation.....</b>	<b>29</b>
Introduction.....	29
Baseline Performance.....	32
Age.....	35
Personality.....	36
Motivation.....	37
Research Questions.....	39
<b>CHAPTER 3: Motivational Features in School-based Cognitive Training.....</b>	<b>44</b>
Introduction.....	44
Methods.....	50
Results.....	56
Discussion.....	63
<b>CHAPTER 4: Cognitive Training and Monetary Compensation.....</b>	<b>70</b>
Introduction.....	70
Methods.....	76
Results and Discussion.....	87
<b>CHAPTER 5: Individual Differences and tDCS-Augmented Cognitive Training.....</b>	<b>105</b>
Introduction.....	105
Methods.....	112
Results.....	116
Discussion.....	123
<b>CHAPTER 6: The Map is Not the Territory.....</b>	<b>135</b>
Conclusion.....	135
Summary of Individual Studies.....	135
The Problem with Binary Questions.....	138
Implications for the Field.....	140

## LIST OF TABLES

<i>Table 3.1</i> Demographic information.....	52
<i>Table 3.2</i> Training improvement by game variant.....	60
<i>Table 4.1</i> Attrition rates.....	77
<i>Table 4.2</i> Baseline trait differences.....	88
<i>Table 4.3</i> Individual transfer measures by group.....	91
<i>Table 5.1</i> Individual-differences regression results.....	118

## LIST OF FIGURES

<i>Figure 1.1</i> Ngram of mentions of “concentration” vs “memory” .....	8
<i>Figure 3.1</i> Object 2-back task.....	53
<i>Figure 3.2</i> Game elements.....	55
<i>Figure 3.3</i> Average game performance on each day....	57
<i>Figure 3.4</i> Average game performance at baseline by condition.....	58
<i>Figure 3.5</i> Training slope by game-type.....	61
<i>Figure 4.1</i> Recruitment Posters.....	78
<i>Figure 4.2</i> Training Improvements.....	89
<i>Figure 4.3</i> Repeated Measures by condition.....	92
<i>Figure 4.4</i> Standardized composite gain.....	92
<i>Figure 4.5</i> Paid group dual n-back training curves.....	95
<i>Figure 5.1</i> Plot of baseline regression results.....	120
<i>Figure 5.2</i> Plot of motivation regression results.....	121
<i>Figure 5.3</i> Follow-up performance.....	123

## ABSTRACT

Despite the potential benefits that improving executive function and working memory might have for performance at school and work, the use of cognitive training as a means to augment these capacities remains controversial. There is consensus that training leads to improvements on the tasks themselves, as well as limited near-transfer. However, there is little agreement regarding whether meaningful far transfer can be demonstrated. In the present work I consider the causes of these varied outcomes, first through a historical perspective, and then through a focus on the influence of individual-difference factors on the outcome of cognitive training. One particular factor that is known to influence the outcome of other interventions, motivation, has not been well-studied in the context of cognitive training. In three studies I examine the influence of motivation, broadly defined, on the outcome of training interventions. Study 1 explores how children might respond to different versions of a cognitive training program that include certain game-like features that are thought to be motivational. Contrary to my expectations, I found that children performed better on the versions of the training that did not include certain common “engagement” features. I interpret this finding as evidence that features that distract from the core task might actually reduce performance early in training. Study 2 examines the influence of extrinsic monetary reward on the outcome of a training program in young adults. Again, contrary to my expectations, I found that payment did not have an undermining effect on training or transfer performance, although it may have been responsible for some differences at baseline. Study 3 explores several individual difference factors, including

motivation and baseline cognitive performance, on the outcome of a transcranial direct current stimulation (tDCS) augmented working memory training program in adults. I found that both baseline performance and motivation influenced the outcome of the intervention, but only among participants who did not receive active stimulation. From these studies I conclude that motivation may have a nuanced and multifaceted influence on the outcome of training interventions. Finally, I briefly discuss the implications of these findings and what might be done to improve future cognitive training research more generally.

## CHAPTER 1

### The Origins of the “American Question”

**Hall:** Now that we’ve mentioned an American educator, may I ask what you have called “the American question”? Is it possible to speed up the learning of conservation concepts?

**Piaget:** In turn may I ask the counter-question? Is it a good thing to accelerate the learning of these concepts?

Elizabeth Hall & Jean Piaget, Interview in *Psychology Today*, 1970

#### Introduction

It is often recounted that, when renowned Swiss psychologist Jean Piaget was asked whether it was possible to accelerate the acquisition of concepts such as conservation of number or volume beyond the typical course of development, he referred to the query as the “American question.” This response was so common that journalist Elizabeth Hall, who developed a considerable reputation for her interviews of prominent psychologists, made it a point to ask it of Piaget in her 1970 interview for *Psychology Today* (Hall, 1970). Piaget’s response was his typical one – that he saw no reason to accelerate what was dictated by nature and already optimal.

And yet, a brief internet search suggests that few have taken Piaget’s response to heart. Advertisements for applications and programs developed by the nascent “brain training industry” abound. Lumosity, Elevate, and Peak, for example, are listed on a popular technology website as “three brain training apps that really work” (Harper, 2016). While the development and sale of

games designed to improve cognitive function has only recently become a billion-dollar industry<sup>1</sup>, these programs are anything but new. For example: the issue of *Psychology Today* immediately before the one that featured the interview with Piaget included an advertisement for a logic game called WFF N'Proof. Just a few years earlier, WFF N'Proof itself was studied as a means of improving performance on the California Test of Mental Maturity (Allen, Allen, & Miller, 1966). It was only one of many historical training interventions essentially designed to answer the “American question.” In fact, the list of studies that examine the effects of experience on cognitive performance is so vast that it defies inclusion in any single review, meta-analysis, book, or dissertation. Here I focus on a single thread of this research: cognitive training. The remainder of the present chapter will serve to define the topic and provide historical perspective on why it has long been a compelling means of attempting to answer Piaget’s “American question.”

### **What is Cognitive Training?**

Cognitive training (or “brain training,” or “mind training”) refers to activities designed to make people “smarter” and thus better at reasoning, problem solving, and learning. Many current cognitive training programs target basic cognitive skills such as *attention* (the ability to selectively attend to relevant information), *working memory* (the ability to actively keep in mind task-relevant thoughts), or *executive functions* (the set of processes involved in controlling and regulating thought and action). The focus on these processes arises from the fact that there are very real limits on their capacity and that individuals differ in terms of these limits (Daneman & Carpenter, 1980), that they are necessary for complex, intelligent behavior (Carpenter, Just, &

---

<sup>1</sup> The “billion dollar” number is often cited from a single industry report from sharpbrains.com, however, the combined sales of all major brain training companies never reached this number during my time in the industry.

Shell, 1990), and that they are highly correlated with individual differences in intelligence, academic achievement, and life outcomes (Engle, Kane, & Tuholski, 1999; Clair-Thompson & Gathercole, 2006). In recent years, many cognitive training programs have utilized tasks that were originally created to help understand these processes (e.g. Jaeggi et al., 2008; Rueda et al., 2005). Given their importance, it is not surprising that researchers have long been interested in their potential for malleability (Kay, 1888). Although questions of how these processes operate are not fully resolved (see for example Miyake & Shah, 1999) researchers know that the prefrontal cortex is the primary brain region associated with them (Braver et al., 1997; Nee, Jonides, & Berman 2007).

Historically, similar, but non-computerized, tasks were used in attempts to enhance attention and memory (Feuerstein & Jensen, 1980). Other interventions embed these basic processes in other activities, such as play (e.g., Tools of the Mind, Diamond et al., 2007). More complex activities that are thought to transfer to skills like reasoning are sometimes incorporated into cognitive training as well (e.g., problem solving). In addition to activities developed specifically to enhance cognition, sometimes off-the-shelf activities (e.g., video games, board games, dance, music) are used for the purpose of improving reasoning and problem solving (Diamond, 2012; Mackey et al., 2011).

It is known that practice on the activities described above leads to improved performance on those same activities, but to what extent do those improvements matter for other, untrained tasks? For some things, it is assumed that practice does transfer to other situations. A basketball player who lifts weights or practices sprinting does so not to improve those basic skills, but to become a better basketball player (these athletic examples are also apt because of the too-frequent assumption that fade out effects mean that training programs have failed, while it may

simply be that, as with physical exercise, continued practice is necessary to reap the benefits). The degree that practice-based improvements transfer to *other cognitive tasks* is, however, a matter of controversy (and has been for some time; see Thomson, 2006; Yellowlees, 1940). Cognitive skill improvements are relevant for a wide range of populations, from older adults whose cognitive capacities might be in decline to fighter pilots who need to perform at peak capacity. Additionally, children who have ADHD, experiences of early stress due to poverty, nutrition deficits, and so forth all might benefit from cognitive interventions. In fact, cognitive training is potentially relevant for everyone, even those whose abilities are within the normal range. Thus it is not surprising that so many people are interested in mind training, and have been for many years. And yet there is often little discussion in present research of the origins of this interest – and furthermore, of the early attempts to study such programs empirically.

### **A Problem of Semantics?**

A key challenge in identifying early cognitive training interventions and research lies in the shifting language used to describe “cognitive performance.” Terms like attention, concentration, reasoning, and memory have long histories within the English language but have generally retained consistent meaning within the vernacular, whereas terms such as “intelligence” have changed dramatically following certain key events (in this case, the development of psychometric testing at the turn of the 20<sup>th</sup> century). Other popular terms, such as “executive function,” have only surfaced much more recently (Baddeley, 1979). Given the popularity of keyword searches in determining what papers to include in reviews and meta-analyses, the issue of terminology may heavily limit the inclusion of historical work in discussions of cognitive training.

For example, there is a large body of research detailing the relationship of experience on *intelligence per se*. While this historical body of work goes largely unmentioned in contemporary reviews (Simons et al., 2016 being the notable exception), the work that is mentioned is nevertheless limited to articles that explicitly use the term intelligence. For example, one can follow the track of intelligence-focused intervention research in scholarly pieces from the 1928 *Yearbook of the National Society for the Study of Education* (Whipple, 1928) to the 1940 *Yearbook* of the same society (Whipple, 1940) to J.M. Hunt's landmark *Intelligence and Experience* (1961) to Whimbey's *Intelligence Can Be Taught* (1975) to Detterman & Sternberg's *How and How Much Can Intelligence Be Increased* (1982) to Spitz's *The Raising of Intelligence* (1986) to Martinez's *Education as the Cultivation of Intelligence* (2000) to Jausovec's *Increasing Intelligence* (2017), which was published this very year and discusses Study 1 of this dissertation. Each of the above pieces are extremely comprehensive and include detailed discussion of most of the major studies that came before it. Some, like Detterman & Sternberg's book, do an admirable job of collecting almost every major study designed to improve intelligence.

And yet, because these works focus on intelligence *per se*, they leave out a variety of interventions and scholarly work aimed at related constructs such as attention or reasoning. For example, despite having his work promoted by William James (James, 1890), very few contemporary researchers are aware of David P. Kay's work (1888), and when they do mention him, it is largely in the context of improving long term memory (Collins, 2014). While this is understandable given that his book is titled *Memory and How to Improve It*, Kay understood and defined mental processes in ways that would not seem out of place in modern discussions of executive function. For example, Kay writes of a "second type of memory" that he refers to as

the “rational” form. “It is of the utmost importance to us, in forming our judgement of things or in determining upon a particular line of conduct, to be able to bring together before the mind a number of instances of the same or a like kind, recent or long past, which may aid us in coming to a right determination” (Kay, 20). This description fits well with modern definitions of working memory. Kay also described attention and inhibitory control, and what might be done to improve them. However, Kay’s book rarely mentions intelligence, and thus it is perhaps unsurprising that it has been left out of historical reviews of interventions to improve cognition. That this distinction often comes down to semantics has not entirely been lost on researchers interested in interventions designed to improve cognitive function. For example, in Brown and Campione’s chapter within *How and How Much Can Intelligence Be Increased*, they argue that most of the interventions detailed earlier in the book are focused more on *cognitive skills* rather than intelligence more broadly (Brown & Campione, 1982). And while they make a compelling case that this is indeed true, they do not consider whether studies that explicitly focus on cognitive skills, and their historical antecedents, might have been left out from their discussions. Thus, in the present chapter, I try to focus on interventions that fit the general definition of cognitive training described above, rather than studies specifically targeted at improving a particular construct, whether that be working memory, reasoning, or intelligence.

### **Historical Interventions in the Public Eye**

Of course, part of the reason that modern researchers see David Kay as a memory researcher is that many of his contemporaries were indeed focused on improving long term memory. Why do most contemporary cognitive training interventions focus on improving cognitive skills, rather than memory per se? Sources from antiquity, reaching as far back as the Dhammapada in the 3<sup>rd</sup> century B.C.E. (Buddhaghosa, 1996), suggest that humans have long

recognized that being able to attend to the world and inhibit distractions are key in a successful mental life and that these capacities may be improved through practice. However, prior to the industrial revolution many documented mental training activities in the West, such as the mnemonics of Simonides and Saint Thomas Aquinas, focused on long-term memory (Patten, 1990), although Plato was cognizant of the idea that training in arithmetic could impact general mental quickness (Grube, 1992).

Scholarly textual references to improving constructs psychologists would now consider “cognitive skills” – such as concentration, attention, and inhibition – generally remain rare until the second half of the 19<sup>th</sup> century. While there are likely a variety of causes for this, a review of early training programs and early interventions suggests a convergence of three specific factors. First, by the 1880s in America and Europe, compulsory education was generally accepted as a public good. A renewed focus on educational pedagogy for the masses cultivated new interest in techniques to help a wider variety of students prepare for the workforce. The resurgence of popularity of Herbart in the mid-1800s is a primary example of this (De Garmo, 1895; Boring, 1950). Second, by the final decades of the 19<sup>th</sup> century, the disciplines of psychology and neuroscience were formalized, both within the minds of the educated public and the academy. Work by William James helped to embed constructs such as *attention* as a topic of study; the early psychometric work of Francis Galton helped to operationalize intellect as something that could be explicitly measured (and thus raising the question of whether those measurements might change over time) (Hergenhahn & Henley, 2013). Finally, by the late 1800s there was a greater public interest in skills that would now be described as related to *learning to learn* rather than on memorization per se. For example, terms related to the non-memory cognitive processes

necessary for learning and working, such as concentration, saw considerably greater mention in media and books during this time (see Figure 1.1).

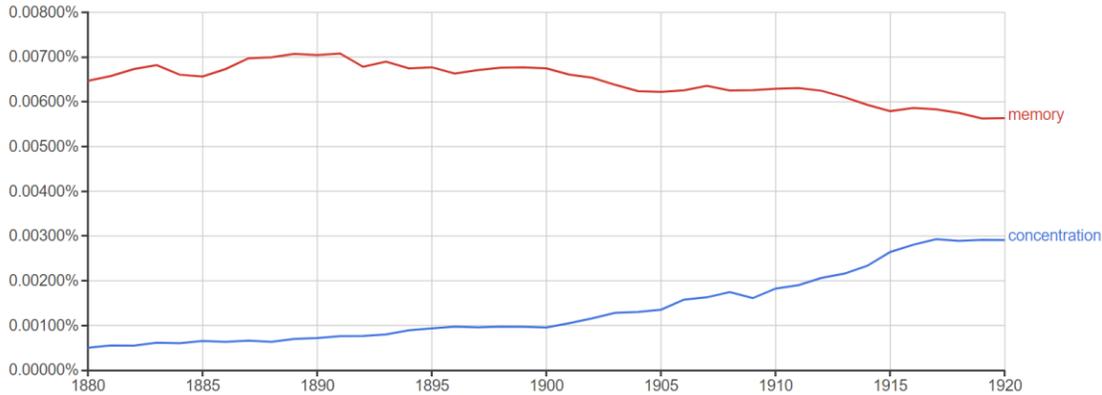


Figure 1.1. Ngram of mentions of “concentration” vs “memory”, late 19<sup>th</sup> to early 20<sup>th</sup> century

At first this meant that mnemonic programs were discarded in favor of other strategy-based systems for memorization (Collins, 2014). The popular memory improvement programs, such as Dr. Edward Pick’s popular *On Memory and the Rational Means of Improving It*, included language that generally depicts improving memory itself as a means to an end (Pick, 1860). Pick describes memory as “the foundation of intellectual life.” However, similar programs of the late 1800s (including that of the pseudonymous “Alphonse Loiset,” who infamously copied Pick’s approach in his own wildly successful program) included marketing language more closely focused on improving “attention” rather than memory alone (Loisette, 1896). Loiset (actually Marcus Dwight Larrowe) operated his own memory improvement clinic at 237 5<sup>th</sup> Avenue in New York, and counted Mark Twain among his clients. While memory was certainly a focus of Loiset’s program, his marketing took a different tack. He wrote that in addition to making “surer remembers” his system “does and must make better observers, clearer and more consecutive thinkers, and sounder reasoners” (Loisette, 1896).

The many spiritualist movements in the United States, and a Victorian area focus on health improvement, including the health of the mind, also brought “mind-training” to the forefront of the public sphere. Pamphlets for programs of this time, such as the *Ralston Brain Regime*, “designed to develop perfect health in the physical brain, strengthen the mind, and increase the power of thought” may have literally been sold alongside snake oil supplements (Ralston, 1891). Another program, Pelmanism, brought “mind-training” to popular consciousness across Great Britain and later much of the world (suggesting that the possibility that cognition might be improved has never been a uniquely American question). It combined self-help invectives with the completion of repeated cognitive tasks; among the activities it included was the card game *Concentration* and the chess-based knight’s journey. Pelmanism was probably the first example of widely-available commercial brain training and at its peak counted over 500,000 customers worldwide (Katz, 2016; Thomson, 2006). By World War II, however, the scientific community firmly saw Pelmanism as lumped in with “autosuggestion... unfired food, dietetic and psychological magic” (Yellowlees, 1940) and other offerings in which “prestige and profits are acquired by dubious interventions in the lives of others” (Sullivan, 1942).

### **The Example of Catherine Aiken**

While the “Pelman Institute” was printing its first little grey books in the 1890s, Catherine Aiken, a Quaker schoolteacher in Stamford, Connecticut, developed a system of attention training for the girls in her school at least a decade earlier. Like many cognitive interventions of today, it required pupils to spend about 15 minutes a day on short attention and memory activities. In 1894, Charles Warner described her system in *Harper’s* (Warner, 1894). In one activity, “a collection of figures was placed upon the reverse side of a revolving blackboard,

then quickly turned; the figures were instantly recognized in their order” (i.e., a working memory task). To make the task more difficult, some exercises required students not only to memorize numbers but also to apply various arithmetic operations to them. Her program also included *subitizing* practice long before Kaufman coined the term; the *Harper’s* account describes it as follows “Another exercise which developed quick perception is that of ‘unconscious counting,’ or of immediately recognizing the number of a group of objects without counting them” (Warner, 1894). Following the release of the Harper’s article, Aiken published two books, *Methods of Mind-Training: Concentrated Attention and Memory* and *Exercises in Mind-Training In Quickness of Perception, Concentrated Attention, and Memory* (Aiken 1895; Aiken, 1899). These books describe her program in detail and provide fascinating accounts of the “action research” Aiken carried out. Aiken was interested in developing more than her students’ attention and memory, however: “This power of concentration has been sought for, not with the idea of making mere memorizers, but in order that they may be able to recall promptly what they have gathered from the great realm of facts and principles, so as to hold it in the mind as a basis if reasoning, and ultimately, for the purpose of possessing well disciplined and self-controlling minds” (Aiken, 1895). In other words, Aiken believed that her program led to successful *transfer*. But even early accounts of Aiken’s work were skeptical of this claim. L.H. Galbreath, of the University of Buffalo, wrote of Aiken’s training program: “However, a great danger for theoretical and practical pedagogy arises out of the assumption of the possibility of training a power to attend to things in general from special formal exercises. Because one acquires special power to attend to things of sight, it does not follow that he can attend with equal skill and efficiency to sensations of sound” (Galbreath, 1897).

Aiken was eager to establish that her work was not associated with “animal magnetism, hypnotism, and other isms” (30). What is truly remarkable about Aiken’s program— and what sets it apart from the “isms” of the time, such as Pelmanism— is the attention it received from psychologists and educational researchers. Aiken’s first book on mind-training, for example, includes an encouraging letter from G. Stanley Hall, the first President of the American Psychological Association. In 1907, G. M. Whipple presented research conducted on Aiken’s program at the American Association for the Advancement of Science. A brief account of this presentation was published in *Science* the following year (Mann, 1908), and a detailed account of the study two years later in *The Journal of Educational Psychology* (Whipple, 1910). To evaluate Aiken’s system under “laboratory conditions” Whipple, then at Cornell, conducted two experiments. Experiment 1 tested 6 college students before and after practice on the letter memorization component of Aiken’s exercises for approximately an hour a day. Experiment 2 included 3 adult participants and a broader range of Aiken’s exercises for 3 hours a week for 7 weeks. Whipple used a tachistoscope to prevent “eye-moving or the roving of attention” rather than Aiken’s revolving chalkboard (Whipple, 1910). Whipple found no evidence that training on these exercises led to any general improvements. Instead, he found “a very slight effect” of practice “which is easily explicable in terms of habituation to the experimental conditions and of development of the “trick” of grouping.” W.S. Foster, a student of Whipple’s, supplemented the original study with what he believed was a significant improvement: each participant was a trained psychologist. Foster arrived at similar results (Foster, 1911), “That training in these experiments has made the observers noticeably better observers or memorizers in general, or given them any habits of observing closely or reporting correctly, or finished any ability to meet

better and situations generally met with, neither we nor any of the observers themselves believe. It seems, therefore as if the value of formal training of our kind had been greatly overestimated.”

Foster and Whipple both noted that their experiments were imperfect, and that the level of participant experience, or the duration of practice and age of the participants may have impacted their results; Whipple himself adds a footnote expressing regret at being unable to conduct his experiment with children. However, he also adds, in reference to the issues above, “neither of these objections seem to us of great moment; we feel that our observers had reached their maximal efficiency, and we are unable to believe that children could be brought to exhibit a range of apprehension so markedly superior to that of competent and well-trained university students and instructors” (Whipple, 1910).

Although Whipple and Foster did not conduct follow-up experiments with children, Karl Dallenbach (a student of Titchener) did. Dallenbach conducted his study within the Ithaca, NY public school district, with 29 students. Students trained for ten minutes daily for 17 weeks, with progressively more difficult material; furthermore, pre-tests, post-tests, and follow-up tests (41 weeks after training) were created with untrained material (Dallenbach, 1914a). Unlike Whipple and Foster, Dallenbach found that his students did improve, particularly those initially classified as having “poor” performance. These improvements persisted at follow-up. Dallenbach collected not only grades of his students (which rose following the intervention), but also performance on an early *Binet Test of Attention*. Furthermore, Dallenbach compared the trained students’ performance to a set of students who had not been trained (although this test was only collected at post-test). Dallenbach noted that grades were significantly higher following the intervention, and that students who completed the training outperformed those who had not done so on the Binet attention test. Like Whipple and Foster before him, Dallenbach was aware of many of his

methodological limitations, but he arrived at very different conclusions. “Our more lengthy experiments with children, however, have not only showed more decided practice-effects, but have also rendered it at least quite possible, if not practically certain, that these practice effects have brought about a permanent modification in the mental traits exercised, and what is more, a modification that certainly seems to have made itself felt in a number of ways outside the special tests we made (as in an improvement in school work and increased efficiency long afterward in supplementary tests of observation and report)” (Dallenbach, 1914b). In 1919 Dallenbach repeated his experiment with children with cognitive deficits and arrived at a similar conclusion (Dallenbach, 1919), a somewhat remarkable effort considering that even some modern researchers improperly generalize training findings between different study populations.

Whipple, Foster, and Dallenbach eventually moved on from this research, but interest in the improvement of basic cognitive skills continued. There are far too many individual studies to list in a single review, but some are briefly detailed here. Some of these were used quite widely, such as Feuerstein’s “Instrumental Enrichment” program: Reuven Feuerstein, working from a Piagetian perspective in the 1960s and reflecting on his experiences with young Holocaust survivors, created a variety of facilitator-administered pen and paper tasks meant to improve memory and attention in school (Feuerstein & Hamburger, 1965). Others were designed for specific purposes, such as the Space Fortress computer game, funded by the Defense Advanced Research Projects Agency (DARPA) with the ultimate goal of improving prefrontal function in highly cognitively demanding jobs (e.g. fighter pilots) (Donchin, 1995). In addition to attempts to improve basic cognitive processes, educators and psychologists tested the potential effectiveness of reasoning training, logic, philosophy, and even Latin language learning to assess whether or not these skills could impact academic achievement and thinking more generally. For

example, the “academic games” designed by Layman Allen in the 1960s, including *WFF N’ Proof* and *Equations*, were the subject of at least two controlled studies (Allen, Allen, & Miller, 1966; Allen, Allen, & Ross, 1970). Programs developed or used outside of English-speaking countries often received less attention. For example, *Project Intelligence*, a reasoning training program developed for classroom use and offered widely in Venezuela, saw minimal adoption in the United States (Nickerson, 1985). Also, the children’s concentration program developed by Kossow and Vehreschild in East Germany in the early 1980s generally goes unmentioned, despite promising findings (Kossow & Vehreschild, 1983; Vehreschild, Kossor, & Schulz-Wulf, 1984). Finally, while many children of the 1980s remember *Logo* as their introduction to programming, many do not realize that an original motivator behind the program was the development of cognitive skills more generally (Clements, 1985). The studies testing these interventions included transfer tests of various kinds, including IQ tests, academic achievement, and complex reasoning. Many of these studies led to critical discourses that recalled the debates regarding Aiken’s intervention; for example, see Stanley and Schild’s 1971 response to Allen’s work (Stanley & Schild, 1971), or Shayer and Beasley’s (1987) discussion of Feuerstein’s Instrumental Enrichment.

Given the wide-ranging extent of these interventions, it may seem that a salient question is what sort of boundary conditions should be place around the definition of cognitive training. For example consider three early programs focused on intervention in early childhood: Project Headstart, The Milwaukee Project, and The Abecedarian Project all included, to some extent, activities that could be considered designed to make children better at “reasoning, problem solving, and learning” (Detterman & Sternberg, 1982). Should they be considered examples of cognitive training? For that matter, what about “normal” educational curricula that includes

components that arguably improve executive function? I pose these questions here but do not suggest that clear answers to them are necessary. Rather, it may be useful to consider this vast landscape of programs when drawing conclusions about “cognitive training” from the outcome of any individual study. For example, should conclusions about cognitive training drawn from an online, computerized, game-like intervention be applied to classroom interventions with a strong social component?

What *can* one conclude from the example of Catherine Aiken and a review of historical work more generally? First, activities that would today be classified as “cognitive training” have long been a subject of heated debate in psychology and education. Second, the focus of the research has, from the very start, been on whether “it” works, rather than *why* “it” might work, under what conditions, and for whom; this is despite the fact that early researchers noted that these factors mattered. Third, even though researchers noted limitations in their methods (sample size, age, amount of training, experimental design, etc.), they nonetheless felt comfortable drawing sweeping conclusions about the effectiveness of cognitive training in general. When scientists began to focus on specific questions about cognitive training for specific populations (e.g., can Space Fortress help military personnel perform complex tasks), they received less attention, possibly because neither scientists nor the media were inappropriately generalizing the findings.

This section may rightly bring to mind the old debate over “formal discipline;” that is, the question of whether training or experience in one area may transfer to another skill or intelligence more generally (see Lehman, Lempert, & Nisbett, 1988 for a more recent study; 18 for an early historical review). This was often used as justification for teaching Latin or math, even if these subjects did not have obvious utility in everyday life. Whipple and Dallenbach were

well-aware that their studies were some of the first direct experimental investigations of this theory and that the issue had not been definitively settled by Thorndike and Woodworth (1901), although it certainly went out of favor as a result of their work. As Jenner states in his 1914 review, “for some years now we have not dared to speak of a general mental power and have ventured to mention formal discipline only to kindred spirits and then in a scarcely audible whisper” (Jenner, 1914). Despite this, the many studies included in reviews from Rugg (1916) and Orata (1941) make it clear that cognitive training research continued unabated after Thorndike’s study. Even if it fell out of favor in certain subfields of psychology and education, the *interest* in cognitive training, from scholars and the public alike, continued in various forms throughout the 20<sup>th</sup> century.

Finally, there is evidence that Whipple remained intensely interested in questions about the efficacy of mind-training and the mechanisms of transfer in the years following his initial study. In his preface to C. P. Wang’s 1916 dissertation (Wang, 1916) on visual sense training in children, he wrote that “Contributions to the experimental study of the transfer of training (formal discipline) scarcely need either apology or introduction in a period when, despite the considerable amount of investigation, so very much remains undetermined with respect to the amount of such transfer and the mechanism by means of which it takes place.” Perhaps it was still on his mind in 1922, when he completed his educational psychology textbook, “Problems in Educational Psychology” (Whipple, 1923). One of the problems in the book reads as follows: “Catherine Aiken describes a series of exercises (columns of figures, groups of dots to be counted, important dates, sets of drawings, etc.) to be placed on a revolving blackboard, which is then whirled about before the pupils in such a way as to expose the material for a few seconds only. These exercises are strongly urged as a means of developing concentrated attention, quick

and accurate observation, and of accelerating the whole process of learning. Miss Aiken reports very wonderful results from the use of such exercises for five or ten minutes daily.” He then asks “Is there psychological warrant for the use of such exercises as a means of developing attention and observation? Would you advocate the introduction of such exercises as a stock feature of school training?” A review of more recent literature illustrates that the same questions that Whipple fixated on remain unanswered today.

### **The Present Issue**

In 2008, Susanne Jaeggi, Martin Buschkuhl, and colleagues published their graduate work in the Proceedings of the National Academy of Sciences (Jaeggi et al., 2008); they found that practice on a dual n-back task led to improvements in fluid intelligence. The dual n-back task requires individuals to listen to a stream of letters and judge whether a letter was the same as the one presented  $n$  trials previously, while simultaneously viewing a set of boxes on a screen and judging whether the same box “lit up”  $n$  trials previously. Fluid intelligence is defined as the ability to solve abstract, novel problems that require little knowledge and was measured before and after training by matrix reasoning tests that require participants to judge which of several options best fits into an array of figures. Tests of fluid intelligence are correlated with working memory and prefrontal function more generally (Engle, Kane, & Tuholski, 1999) because they require keeping track of and testing numerous rules during the course of problem solving (Carpenter, Just, & Shell, 1990).

The dramatic improvements detailed by Jaeggi et al. (2008) received a considerable amount of attention from the scientific community and the popular press. Additionally, companies offering cognitive training software often took advantage of their findings for marketing purposes (e.g., Learning RX, Lumos Labs, and CogMed). The media hype around

Jaeggi's paper emphasized its putative novelty; for example, Alexis Madrigal (2009) wrote in *Wired*, "Fluid intelligence was previously thought to be genetically hard-wired." As the historical summary above suggests, such claims were inaccurate. Many studies explicitly tested and found improvements in fluid intelligence, even if they did not necessarily use the term fluid intelligence. More generally, there has always been a debate regarding the relative importance of nature (i.e., genetics) and nurture (i.e., experiences) in the development of intelligence. One would be hard pressed to find a scientist who argues that intelligence (including fluid intelligence) is entirely genetically determined and not at all affected by experience. The Jaeggi et al. (2008) paper is a bit more nuanced in discussing this issue than the media reports and acknowledges that there is a history of cognitive training research but that successful transfer has been difficult to achieve.

The interpretation of the Jaeggi et al. (2008) study in terms of a paradigmatic shift within a false dichotomy of fixed versus malleable intelligence, with little attention to historical context, is one reason for the swift critique the study received. There were also numerous concerns regarding the research methods of the study, most notably, the lack of an active control group. Furthermore, scientists were concerned that the general public might expend resources on unproven products, possibly to the detriment of other beneficial activities. Adding to the controversy, Redick and colleagues tried to replicate the Jaeggi findings with a somewhat better controlled trial but found no evidence of gains in fluid intelligence (Klauer & Phye, 2008). Many additional studies continued to ask the question, "does cognitive training improve intelligence?" One could replace "intelligence" with attention or working memory, and find an equally generous body of work. However, this question, like the question "does medicine cure

disease,” is inappropriate. That it continues to be asked, over 100 years after the studies of Whipple and Dallenbach, should give researchers reason to pause and take stock.

Why do some studies find a positive impact of cognitive training whereas others do not? One reason is that “cognitive training” refers to such a broad range of activities (e.g., commercial programs like *Cogmed*, laboratory tasks such as the n-back, and off-the-shelf games). It is not possible to draw conclusions regarding cognitive training as a whole with a single empirical study. The extent to which one can reasonably generalize from one intervention to others is not clear, and researchers are not yet well-aware of what intervention characteristics may be important for transfer. Consider, for example, different working memory interventions. In addition to the n-back task, one can train working memory by having individuals remember sequences of items (*i.e.*, span tasks; see (Chein & Morrison, 2010). Training might be spaced across time or take place within a shorter time frame (Wang, Zhou, & Shah, 2014. And studies may involve fixed block of training (say, across one month) or add “booster” sessions later on (Ball et al., 2002). Cognitive interventions may vary on numerous other dimensions (which processes are practiced, type of instructions, game-like features, amount of training, computerized or not, etc.).

It is also difficult to judge whether or not interventions are effective above and beyond the influence of various confounding factors. Consider, for example, a study that tests whether improvements on an intelligence test are due to a placebo effect by asking participants about their beliefs. Unfortunately, this too is problematic. Hundreds of participants may be required to adequately test whether or not a construct with a true moderate effect size had an impact on an outcome variable above and beyond a reasonably reliable confound (Westfall & Yarkoni,

2016). Attempting to statistically control for several factors may require impractically large sample sizes.

A related problem is that many studies test participants on a large number of laboratory tasks or surveys but lack the sample size needed to conduct multiple comparison corrections. Our own studies suffer from this concern, as do many others. One concern associated with having a large number of transfer measures or a large sample size has to do with the quality of testing implementation. Outcome measures, when presented to participants in rapid succession, shortened for time constraints, and administered over several hours, may be less reliable than ideal. The Redick et al. (2013) study may have exactly this problem: participants performed seventeen demanding cognitive tests, several of which were shortened. Though the reliability of their tasks is normally high in standard administration, reliability under these conditions is not clear. In general, one ironic aspect of cognitive training research is that a large sample size is crucial, but studies with large sample size have their own problems. Studies with large sample sizes often have much less control over the training regimens or quality of data collection. The Owen et al. (Owen et al., 2010) study, which included thousands of participants, is one such example; the administration of tasks is not at all standardized and training dosage was highly variable.

A related concern is presentation of post-hoc or selective analyses (Schwaighofer, Fischer, & Buhner, 2015). In a follow-up study published in PNAS, for example, Jaeggi et al. tested children on a battery of tests and compared performance of a group that received a single n-back training with a control group that learned science facts (Jaeggi et al., 2011). Overall, there was no impact of the cognitive training intervention on the included measures of fluid intelligence. However, upon noting vast individual differences in improvement on the n-back

task, Jaeggi and colleagues also tested whether or not children who actually improved in the training also improved on matrix reasoning. They did find improvements for this group. Also, children who viewed the training as “too difficult” did not get better on the training task. Jaeggi and colleagues interpreted these findings to mean that some students were easily discouraged and thus did not benefit from the intervention. But this finding could also be explained by assuming that people who can learn well improve from their experiences during training and are also more likely to benefit from taking the same test twice. This is a valid alternative explanation.

One final limitation is that there is minimal testing for ‘real-life’ outcomes (e.g., how much better does a child do in school?). Instead, most outcome measures are laboratory tasks, surveys, or standardized tests. Many studies use performance on matrix reasoning tests as their main outcome measure. Although performance on such tests is correlated with real-life success (e.g., the ability to learn new facts), scoring better on these tests does not mean that one will actually be better in real-world tasks. Some studies do include some real-life outcomes or ecologically valid tasks (e.g., Golden et al., 2014; Deveau, Ozer, & Seitz, 2014), but these studies are few.

Finally, studies also differ in terms of the samples tested. Recall two studies mentioned earlier: Jaeggi et al. (2008) used students from the University of Bern, Switzerland, and found successful transfer to fluid intelligence. Redick et al. (Redick et al., 2013) used students from Michigan State, Georgia Tech, and non-students from the Atlanta area and did not find transfer. There are other methodological merits and concerns regarding both studies, but the populations examined in each study are different enough that it is possible that the divergent outcomes could be driven by demographics. For example, two factors that may influence whether one benefits from training are socioeconomic status and motivation (Segretin et al., 2014; Jaeggi et al., 2011).

The list goes on – personality, age, baseline ability, and many others. But many studies do not examine these characteristics, and too few researchers take the step that Dallenbach did early on to replicate his training study with children who have cognitive difficulties. Thus it is not possible to discern the extent to which these difficulties influence performance (Strobach & Karbach, 2016). Until recently, few researchers focused specifically on the role that individual differences may play in the outcome of these interventions. Nonetheless, Elizabeth Hall, in her response to Piaget’s dismissal of the “American question” in the interview that opens this chapter, does ask whether an individual-difference factor (in this case, writing speed) might in fact provide some support for intervention. While Piaget continued to caution in this interview against any “acceleration” of learning despite this rejoinder, other researchers, such as Richard Snow, have long suggested that the existence of any individual differences in “personality, ability, or motivational characteristics can serve as a source of aptitude or inaptitude for learning from instruction in some given setting, and can thus be a focus for research” (Snow, 1977). Snow, who saw intelligence as “both an aptitude and an outcome for education” (Snow, 1982) would likely not be surprised that the last few years have seen an expanded focus on the role these factors play in the outcome of training interventions. The following chapter focuses more closely on this research.

## References

- Aiken, C. (1895). *Methods of mind-training: Concentrated attention and memory*. American Book Company.
- Aiken, C. (1899). *Exercises in mind-training: In quickness of perception, concentrated attention and Memory*. Harper & Brothers.
- Allen, L. E., Allen, R. W., & Miller, J. C. (1966). Programed games and the learning of problem-solving skills: The WFF'N PROOF example. *The Journal of Educational Research*, 60(1), 22-26.
- Allen, L. E., Allen, R. W., & Ross, J. (1970) The virtues of nonsimulation games. *Simulation Gaming* 1(3), 319–326.
- Baddeley, A. D. (1979). Working memory and reading. In *Processing of visible language* (pp. 355-370). Springer US.
- Ball, K., Berch, D. B. Helmers, K. F, Jobe, J.B, Leveck, M.D., Marsiske, M., et al.(2002). Effects of cognitive training interventions with older adults: A randomized controlled trial. *JAMA*, 288(18), 2271–2281.
- Böring, E. G. (1950). *A history of experimental psychology*. New York: Appleton-Century.
- 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.
- Brown, A. L., & Campione, J. C. (1982). Modifying intelligence or modifying cognitive skills: More than a semantic quibble? In D. K. Detterman & R. J. Sternberg (Eds.), *How and how much can intelligence be increased?* Norwood, NJ: ABLIX.
- Buddhaghosa (1996). *A treasury of buddhist stories. From the Dhammapada Commentar*. Kandy, Sri Lanka: Buddhist Publication Society
- Carpenter, P. A., Just, M. A., & Shell, P. (1990). What one intelligence test measures: a theoretical account of the processing in the Raven Progressive Matrices Test. *Psychological Review*, 97(3), 404.
- 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 and Review*, 17(2), 193–199.
- Clements, D. (1985). Research on logo in education. *Computers in the Schools*, 2(2–3), 55–71.
- Collins, A. F. (2014). Advice for improving memory: exercising, strengthening, and cultivating natural memory, 1860–1910. *Journal of the History of the Behavioral Sciences*, 50(1), 37-57.

- Dallenbach, K. M. (1914). The effect of practice upon visual apprehension in school children. Part I. *Journal of Educational Psychology*, 5(6), 321–334.
- Dallenbach, K. M. (1914). The effect of practice upon visual apprehension in school children. Part II. *Journal of Educational Psychology*, 5(7), 387–404.
- Dallenbach, K. M. (1919). The effect of practice upon visual apprehension in the feeble-minded. *Journal of Educational Psychology*, 10(2), 61–82.
- Donchin, E. (1995). Video games as research tools: The Space Fortress game. *Behavioral Research Methods*, 27(2), 217–223.
- Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, 19(4), 450-466.
- Degarmo, C. (1896) Herbart and Herbartians. New York: Charles Scribner's Sons
- Detterman, D. K., & Sternberg, R. J. (1982). How and how much can intelligence be increased? Norwood, NJ: Ablex
- Deveau, J., Ozer, D. J., & Seitz, A. R. (2014). Improved vision and on-field performance in baseball through perceptual learning. *Current Biology*, 24(4), R146-147.
- Diamond, A. (2012). Activities and programs that improve children's executive functions. *Current Directions in Psychological Science*, 21(5), 335-341.
- Diamond, A., Barnett, W. S., Thomas, J., & Munro, S. (2007). Preschool program improves cognitive control. *Science*, 318(5855), 1387.
- Engel, R. W., Kane, M., & Tuholski, S. (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). New York: Cambridge Univ. Press
- Feuerstein, R., & Jensen, M. R. (1980). Instrumental enrichment: Theoretical basis, goals, and instruments. In *The Educational Forum*, 44(4), 401-423.
- Galbreath, L. H. (1897). Review of Methods of Mind Training. *The School Review*, 5(7), 484–486.
- Foster, W. S. (1911). The effect of practice upon visualizing and upon the reproduction of visual impressions. *Journal of Educational Psychology*, 2(1), 11–22.
- Feuerstein, R., & Hamburger, M. (1965). Proposal to study the process of redevelopment in several groups of deprived early adolescents in both residential and non-residential settings. Available at: <http://eric.ed.gov/?id=ED010780> [Accessed August 19, 2016].

- Goldin, A. P., Hermida, M. J., Shalom, D. E., Costa, M. E., Lopez-Rosenfeld, M., Segretin, M. S., et al. (2014). Far transfer to language and math of a short software-based gaming intervention. *Proceedings of the National Academy of Sciences U S A*, *111*(17), 6443–6448.
- Grube, G. M. A. (1992). *Plato: The Republic*. Indianapolis, Indiana: Hackett
- Hall, E (1970). A conversation with Jean Piaget and Barbel Inhelder. *Psychology Today*, *3*, 25-32.
- Harper, E (2016). 3 brain training apps that really work. Available at <http://www.techlicious.com/guide/brain-training-apps-that-really-work>. Accessed May 15<sup>th</sup>, 2017
- Hergenhahn, B. R., & Henley, T. (2013). *An introduction to the history of psychology* (7th ed.). Boston, MA: Wadsworth
- Hunt, J. (1961). *Intelligence and experience*. New York: Ronald Press
- Jaeggi, S. M., Buschkuhl, M., Jonides, J., & Perrig, W. J. (2008). Improving fluid intelligence with training on working memory. *Proceedings of the National Academy of Sciences*, *105*(19), 6829-6833.
- Jaeggi, S. M., Buschkuhl, M., Jonides, J., & Shah, P. (2011). Short- and long-term benefits of cognitive training. *Proceedings of the National Academy of Sciences of the USA*, *108*(25), 10081–10086.
- James, W. (1890). *The Principles of Psychology*. New York: Holt and Company
- Jausovec, N., & Pahor, A. (2017). *Increasing Intelligence*. London, England: Academic Press
- Jenner, W. A. (1914). The Modern Psychology and Formal Discipline. *The Classical Weekly*, *8*(4), 26-29.
- Katz, B. (2014) Brain-training isn't just a modern phenomenon, the Edwardians were also fans. The Conversation. Available at [theconversation.com/brain-training-isnt-just-a-modern-phenomenon-the-edwardians-were-also-fans-29515](http://theconversation.com/brain-training-isnt-just-a-modern-phenomenon-the-edwardians-were-also-fans-29515). Accessed August 19, 2016
- Kay, D. (1888) *Memory: What It Is and How to Improve It*. K. Paul, Trench.
- Klauer, K.J., & Phye, G. D. (2008). Inductive Reasoning: A Training Approach. *Review of Education Research*, *78*(1), 85–123.
- Kossow, H. J., & Vehreschild, T. (1983). Formation of a concentration training program for concentration disordered children. *Psychiatry, Neurology, and Medical Psychology*, *35*(1), 31–36.
- Loisette, A. (1896). *Assimilative memory, or, how to attend and never forget*. New York: J.J. Little & Co.

- Lehman, D.R., Lempert, R.O., & Nisbett, R.E. (1988). The effects of graduate training on reasoning: Formal discipline and thinking about everyday-life events. *American Psychologist*, 43(6), 431–442.
- Mann, C. R. (1908). Section L, Education. *Science*, 27(692), 525–528.
- Mackey, A. P., Hill, S. S., Stone, S. I., & Bunge, S. A. (2011). Differential effects of reasoning and speed training in children. *Developmental science*, 14(3), 582-590.
- Madrigal, A. (2008). Forget brain age: Researchers develop software that makes you smarter. *WIRED*. Available at: [http://archive.wired.com/science/discoveries/news/2008/04/smart\\_software](http://archive.wired.com/science/discoveries/news/2008/04/smart_software). Accessed August 19, 2016.
- Mann, C. R. (1908). Section L, Education. *Science*, 27(692), 525–528.
- Martinez, M. E. (2000). Education as the cultivation of intelligence. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Miyake, A., & Shah, P. (1999). *Models of working memory: Mechanisms of active maintenance and executive control*. Cambridge University Press.
- Nee, D. E., Jonides, J., & Berman, M. G. (2007). Neural mechanisms of proactive interference-resolution. *Neuroimage*, 38(4), 740-751.
- Nickerson, R. S. (1985). Project intelligence: An account and some reflections. *Special Services in the Schools*, 3(1–2), 83–102.
- Orata, D. P. T. (1941). Recent research studies on transfer of training with implications for the curriculum, guidance, and personnel work. *Journal of Education Research*, 35(2), 81–101.
- Patten, B. M. (1990). The history of memory arts. *Neurology*, 40, 346–352.
- Pick, E. (1866). *On memory and the rational means of improving it* (4th ed.). London: Turner & Co
- Redick, T.S., Shipstead, Z., Harrison, T. L., Hicks, K. L., Fried, D. E., Hambrick, D. Z. (2013) No evidence of intelligence improvement after working memory training: A randomized, placebo-controlled study. *Journal of Experimental Psychology: General*, 142(2), 359–379.
- Rueda, M. R., Rothbart, M. K., McCandliss, B. D., Saccomanno, L., & Posner, M. I. (2005). Training, maturation, and genetic influences on the development of executive attention. *Proceedings of the National Academy of Sciences of the USA*, 102(41), 14931-14936.
- Rugg, H. O. (1916). *The experimental determination of mental discipline in school studies* (No. 17). Baltimore: Warwick & York, Incorporated.

- Schwaighofer, M., Fischer, F., & Böhner, M. (2015). Does working memory training transfer? A meta-analysis including training conditions as moderators. *Educational Psychologist*, 50(2), 138–166.
- Segretin, M. S., Lipina, S. J., Hermida, M. J., Sheffield, T. D., Nelson, J. M., Espy, K. A., Colombo, J. A. (2014). Predictors of cognitive enhancement after training in preschoolers from diverse socioeconomic backgrounds. *Frontiers in Developmental Psychology*, 5, 205
- Shayer, M., & Beasley, F. (1987). Does instrumental enrichment work? *Br Educ Res J*, 13(2), 101–119.
- Simons, D. J., Boot, W. R., Charness, N., Gathercole, S. E., Chabris, C. F., Hambrick, D. Z., & Stine-Morrow, E. A. (2016). Do “brain-training” programs work?. *Psychological Science in the Public Interest*, 17(3), 103-186.
- Spitz, H. H. (1986). *The raising of intelligence: A selected history of attempts to raise retarded intelligence*. Hillsdale, NJ: Erlbaum.
- Stanley, J. C., & Schild, E. O. (1971). Do nonsimulation games raise IQ's, whereas simulation games do not? *Simulation Gaming*, 2(1), 119–123.
- St Clair-Thompson, H. L., & Gathercole, S. E. (2006). Executive functions and achievements in school: Shifting, updating, inhibition, and working memory. *The Quarterly Journal of Experimental Psychology*, 59(4), 745-759.
- Strobach T., & Karbach J. (2016). *Cognitive training: An overview of features and applications*. Berlin: Springer-Verlag.
- Sullivan, H. S. (1942). Psychiatry, the army and the war. *Psychiatry*, 5(3), 403–410.
- The Ralston Brain Regime: Presenting a Course of Conduct, Exercises and Study, Designed to Develop Perfect Health in the Physical Brain, Strengthen the Mind, and Increase the Power of Thought: A Book of Practice, More Than Theory. Available at <https://collections.nlm.nih.gov/catalog/nlm:nlmuid-66250060R-bk>. Accessed August 19, 2016.
- Thomson, M. (2006). *Psychological subjects: Identity, culture and health in twentieth century Britain*. Oxford: Oxford University Press
- Wang, C. P. (1916). *The general value of visual sense training in children*. Baltimore: Warwick & York.
- Wang, Z., Zhou, R., & Shah, P. (2014). Spaced cognitive training promotes training transfer. *Frontiers in Human Neuroscience*, 8, 217.
- Warner, C. (1895). December 1894 to May 1895, Harper's New Monthly Magazine (Harper & Brothers Publishers, New York), Vol 90. Available at <https://archive.org/details/harpersnew090various>. Accessed August 19, 2016.

- Whimbey, A. (1980). *Intelligence can be taught*. New York: Dutton.
- Whipple, G. M. (1928). The transfer of training. *Yearbook of the National Society for the Study of Education*, 27, 179-209
- Whipple, G. M. (1940). Intelligence: its nature and nurture. *Yearbook of the National Society for the Study of Education*, 39, 99-141
- Whipple, G. M. (1923). *Problems in educational psychology: Eighty exercises designed to supply students of education and teachers in training with material for written reports and classroom discussion*. Public school Publishing Company.
- Woodworth, R. S., & Thorndike, E. L. (1901). The influence of improvement in one mental function upon the efficiency of other functions.(I). *Psychological Review*, 8(3), 247.
- Westfall, J., & Yarkoni, T. (2016). Statistically controlling for confounding constructs is harder than you think. *PLoS One*, 11(3), e0152719
- Vehreschild, T., Kossow, H.J., & Schulz-Wulf, G. (1984). Results of concentration training in normally intelligent students with poor concentration. *Psychiatry, Neurology, and Medical Psychology*, 36(3), 152–160.
- Yellowlees, H. (1940). The problem of adolescence. *Lancet*, 6, 233-279.

## **CHAPTER 2**

### **A Map of the Problem: Individual Differences and Motivation**

#### **Introduction**

The list of issues discussed at the conclusion of the previous chapter is not exhaustive, but is intended to provide the reader with some idea for why most studies are far from conclusive. Although it may be easy to scoff at the tiny samples and limited methods used by Whipple in his century-old cognitive training experiments, contemporary studies often share similar issues. Why would psychologists design studies that are underpowered or that have clear methodological problems? In part, they do so because there is a tradeoff such that avoiding one problem (e.g., sample size) leads to another problem (e.g., poor control over intervention). Researchers include a variety of transfer measures all designed to answer different questions: to see if there is change in fluid intelligence measures, academic achievement measures, or assessments of basic skills that underlie more complex measures. It is practically impossible, because of cost and time constraints, to recruit enough participants to make up for the large number of planned statistical tests.

One potential answer to this conundrum is the use of meta-analytic techniques, as these allow researchers to determine not only whether an intervention has a meaningful effect over a large body of studies, but also to reveal potential moderators—such as demographic makeup, pre-existing individual differences, or training dosage—that may influence the outcome of the

intervention. Unfortunately, the extant meta-analyses arrive at very different conclusions and do little to settle the issue (Au et al., 2015; Karbach & Verhaeghen, 2014; Schwaighofer, Fischer, & Buhner, 2015; Melby-Lervag & Hulme, 2013; Melby-Lervag, Redick, & Hulme, 2016; Weicker, Villringer, & Thone-Otto, 2016). While Au et al., Karr et al., and Karbach and Verhaeghen conclude that training executive functions like working memory may be effective in improving capacities such as fluid intelligence, Melby-Lervag and Hulme suggest that transfer gains are non-significant or minor at best. These varied outcomes arise because of key differences in how they were conducted, such as the populations included and the type of intervention used. These decisions, along with the choice of statistical procedures, have a substantial impact on the outcomes of meta-analyses. A nice demonstration of this point is a pair of analyses conducted by Van Elk et al. (Van Elk et al., 2015) about the effect of religious priming on prosocial behavior. Responding to a meta-analysis that found that religious priming has a positive impact on prosocial behavior in religious participants (Shariff et al., 2016), van Elk conducted two publication bias correction analyses (PET-PEESE and Bayesian) using the same data as Shariff et al. Although each of these methods is reasonable, they ultimately arrive at different conclusions. Furthermore, to return to the medication analogy, it is impossible to draw conclusions about a broad question (does medication work) by combining studies of different medications and illnesses in a single analysis. The conundrum is that each individual study differs on so many dimensions that statistically accounting for these differences may be, in essence, the equivalent of reducing the sample size back to the level of individual studies. And as Stegenga (Stegenga, 2011) writes, “Meta-analysis fails to provide objective grounds for intersubjective assessments of hypotheses because numerous decisions must be made when

performing a meta-analysis which allow wide latitude for subjective idiosyncrasies to influence its outcome.”

Thus the existence of meta-analysis as a technique does not end the need to explore factors that may determine the outcome of cognitive training. It is likely that certain individual-difference factors such as age, baseline performance, socioeconomic status, personality, experience with games, and motivation, among many others, may impact the outcome of the intervention for any individual participant. These differences have significant implications not only for our ability to improve our theoretical understanding of cognitive training, but also for the real-world efficacy of any individual intervention. However, many extant cognitive training studies do not examine these factors. Furthermore, most use sample sizes that are too small to adequately account for them individually, let alone the extent they may interact with each other. Conducting larger, better-powered studies that allow scientists to understand the effects of these differences may help to explain the inconsistency across studies thus far. This chapter largely focuses on the evidence that certain individual-difference factors may influence the outcome of cognitive training.

### **Individual-Difference Factors That May Influence Cognitive Training Outcomes**

A full discussion of all the individual-difference factors that *might* influence the outcome of cognitive training would require a book by itself. For a comprehensive review of how individual-difference factors impact cognition more generally, the *Handbook of Individual Differences in Cognition* (Gruszka et al. 2012) provides a detailed discussion. This introduction, however, focuses on individual-difference factors that have been examined in previous cognitive training research. This list, perhaps unsurprisingly considering the previous paragraphs, is fairly short – only a handful of the many cognitive training studies published thus far have explicitly examined individual-difference factors. At present, these studies consider age, baseline

performance, personality factors, and motivation. Readers well-versed in the study of individual differences may notice other factors that may be meaningful predictors in other contexts – such as gender or cultural factors<sup>2</sup> – missing from this chapter. While these certainly merit further study, they are not covered here simply because extant research does not yet suggest that they make a significant contribution to the outcome of a cognitive training paradigm.

### **Baseline Performance**

The potential contribution of baseline performance (either on the training task itself or on the set of cognitive tests used at pre-test) to improvements on untrained assessments merits primacy in a discussion of individual differences. While many individual-difference factors have not been specifically studied in the context of cognitive training, most of these *have* been examined in the context of baseline performance on a variety of cognitive abilities, such as working memory or executive function. If baseline performance impacts the outcome of a training intervention, it is reasonable to suggest that the other individual-difference factors that influence baseline performance merit further investigation; the influence of baseline performance in a domain on its trainability has long been a focus of cognitive training research (Verhaeghen et al. 1992; Willis 1989; see also Snow 1991). Baseline performance on working memory tasks may influence the outcome of cognitive training in one of two directions. One possibility is that those who perform worse prior to the intervention have more room to improve following training, and thus may experience greater gains. Alternatively, those with higher baseline performance may be better able to benefit from completing a cognitive training regimen – they may perform better at the training task over the course of the intervention and thus also experience greater improvements; these

---

<sup>2</sup> While Au et al. (2014) did find differences between cognitive training studies conducted in the United States compared to other countries, it remains unclear whether those differences are related to cultural or other factors (e.g. motivation).

participants may also be more likely to complete the entire intervention and not drop out of a study (Jaeggi et al. 2014). One factor to keep in mind is that the source of individual differences in baseline performance may differ across studies: in some studies, lower baseline individuals may have less experience, be younger or older, and so forth. Thus it is not surprising that baseline performance may have different effects across studies. It is also possible that different training paradigms may result in different patterns of performance across high- and low-baseline participants. For example, process-based training often results in higher gains for individuals with lower baseline performance, while strategy-based training programs often result in greater gains for high-baseline individuals (Karbach and Unger 2014). Thus a sensible approach to resolving these issues is to focus on the underlying individual differences that may influence baseline performance as well as training paradigm characteristics.

Of the few studies examining baseline performance and cognitive training, a number support the former possibility – that is, that those who start with lower levels of performance experience greater gains. In two studies by Zinke et al. (2012; 2014) individuals who performed worse at baseline, across multiple training paradigms of both working memory and executive control, experienced larger gains on the training task. Although Zinke et al. did not directly examine how baseline performance on untrained WM or fluid intelligence measures impacts transfer gains, they did find that, like Jaeggi et al. (2011), Schmiedek et al. (2010), and Chein and Morrison (2010), the amount of improvement on the training task does contribute to the amount of transfer gains on certain executive control and verbal WM tasks. These studies suggest that individuals who begin with lower baseline performance on the trained tasks may have stood to improve more at both the training and related transfer tasks. One possibility, as Zinke and colleagues discuss, is that individuals with higher baseline performance may be closer to ceiling

performance at the task (some tasks are designed with a fixed level of maximum challenge, while others remain adaptive as long as participants continue to improve). If improvement in the task is a necessary precursor to transfer gains, it is also possible that modifying the task to permit high-performers to continue improving beyond present ceiling levels might also permit them to experience greater transfer gain.

Few studies have specifically looked at how pre-test performance on the transfer tasks may influence transfer gain, although consistent with the previously mentioned small studies, one recent, large-scale study found that individuals who performed worse at pre-test on the set of transfer tasks also showed greater improvements on these tasks following training than those with higher pre-test scores (Hardy et al. 2015). This finding is also consistent with research conducted on the ACTIVE training project with older adults (Willis and Caskie 2013) that found that lower performance on certain baseline measures was correlated with greater improvement after a period of cognitive training. While these studies provide some evidence that lower-performing individuals may stand to benefit more from the training than those who are closer to ceiling, the relationship between baseline ability and transfer may be fairly complex and might also be influenced by methodological differences, such as the design of the intervention or the adaptivity algorithms used to increase or decrease the difficulty of training. And finally, some of the outcome measures might not be sensitive enough to detect changes at the upper end of the scale<sup>3</sup>, which could also contribute to the relatively smaller improvements observed in high-ability samples.

---

<sup>3</sup> For example, the difficulty of the adaptive n-back task does not necessarily increase in a linear fashion; the increase in ability necessary to advance may be higher from 5- to 6-back than from 2- to 3-back. However, it is difficult to measure improvements in ability in participants at higher n-back levels unless they result in a new n-back level being reached.

## Age

A substantial body of research provides evidence for the effects of age on cognitive plasticity across the lifespan (Guye et al., 2016); it should be unsurprising that age has often been linked to differences in transfer improvements following cognitive training. Several studies have found that improvements on untrained tasks are smaller for older adults than younger adults (Zinke et al. 2014; Brehmer et al. 2012; Schmiedek et al. 2010), and even smaller for old-old adults, when compared to young-old individuals (Borella et al. 2014). However, meta-analytic work has revealed inconsistent findings on this issue: While one recent meta-analysis found no difference between younger and older adults in transfer improvements, (Karbach and Verhaeghen 2014), another found that younger adults improved more on these tasks than older adults (Wass et al. 2012). Considering the fact that that these meta-analyses include different sets of studies based on differing parameters (for example, Wass et al. include a larger range of ages than Karbach and Verhaeghen), it is difficult to compare them to each other.

Given the extent to which age impacts baseline performance on a wide variety of cognitive tasks (Salthouse 1996), there is a reasonable impetus for examining the effects of age in cognitive training research. Furthermore, if training mitigates the effects of age-related cognitive decline, older adult populations may benefit the most from cognitive training (Richmond et al. 2011). Perhaps most problematic, from the perspective of critiquing extant research, is that age effects have generally been examined only in the context of older versus younger adults, with the exception of Borella et al. (2014) mentioned above. Surprisingly, little is known about how age may impact transfer (or study completion, for that matter, see *Motivation* below) in individuals who are older than college age but younger than retirement, or among children of varying ages. Until a truly comprehensive study is made that includes the entire lifespan – from children to young

adults to middle age to older adults – a significant gap remains in our understanding of age as an individual-difference factor relevant to cognitive training research.

### **Personality**

Since most cognitive training work has been conducted by cognitive psychologists, it is perhaps unsurprising that personality and temperament have not been thoroughly examined in the context of cognitive training. However, some recent research suggests certain parts of the five-factor personality model may be predictive in the outcome of these interventions.

For example, Studer-Luethi and colleagues (2012) found that individuals who had high ratings on the conscientiousness portion of the five-factor inventory were more likely to perform well on both an n-back working memory training intervention as well as related near transfer tasks. While this finding makes sense given that conscientiousness is generally associated with persistence and self-discipline, higher conscientiousness was associated with *lower* far transfer performance. The authors suggest that the participants developed task-specific skills that prevented far transfer. This result merits further investigation, particularly if one takes into consideration the fact that this study was fairly well-powered.

In the same study, Studer-Luethi et al. (2012) found that that higher levels of neuroticism were associated with significantly higher levels of transfer to matrix reasoning tests for participants who completed a single n-back version of the training, but that lower levels of neuroticism were associated with higher transfer scores for individuals from the dual n-back group. Studer-Luethi and colleagues posit that the dual n-back version of the task is more difficult than the single n-back version, and this “led subjects with high levels of neuroticism in a suboptimal activation state which derailed complex cognitive transfer processes” (2012).

However, this explanation is not completely satisfactory, given that participants in both conditions trained on an adaptive version of the n-back task and arguably should have been similarly challenged, regardless if a single or dual n-back version was used. Since there was no significant main effect of neuroticism within both groups combined, another possible explanation is that neuroticism does not play an important role in the outcome of training. Similarly, in another reasonably well-powered study, Urbánek and Marček (2015) find a differential relationship between training type, personality, and transfer such that participants in the single n-back training group (but not a dual n-back or mental rotation condition) experienced greater transfer gain if they rated higher on traits generally associated with neuroticism. Again, however, it is worth noting that this study found no significant effect of training on transfer versus an active control group, and that this personality finding was only revealed after examining a training group on its own. Thus, an argument could be made that the (limited) extant research does not suggest that personality plays a consistent role in the outcome of cognitive training.

### **Motivation**

Despite being an entire subfield of psychological research, little research has examined how a participant's motivation, either to complete a training intervention or to improve his or her cognitive capacity, impacts training and transfer. Few studies examine the impact that self-efficacy, goal orientation, mastery beliefs, mindset, and many other motivation-related constructs. There is some evidence, however, that this is a space worth exploring. For example, a number of previous training studies inform participants that they may improve their intelligence or cognitive function during the study (for example, Jaeggi et al. 2008; Klingberg et al. 2005); while other studies only mention practicing computerized tasks (Redick et al. 2013). One study suggests that personal beliefs about the malleability of intelligence may contribute to the amount of transfer

after a cognitive training intervention (Jaeggi et al. 2014). Individuals who believed that intelligence could be improved experienced larger transfer gains following training. Although the beliefs-by-intervention interaction was not significant in this instance, it does suggest that personal beliefs about whether one is able to improve cognition—itsself a major factor in how motivated one might be to complete cognitive training—could have a substantial impact on the outcome of training. Similarly, a study by Foroughi and colleagues (2016) found that participants recruited with materials that promoted IQ gains led to greater improvements after a training intervention than those recruited through flyers without this messaging. However, this study only included a single training session. Nonetheless, these studies provide preliminary evidence that Dweck’s mindset work might have relevance in helping to understand the outcome of cognitive training (Dweck, 1999). This perspective could explain the outcome of Foroughi’s study, for example: if the recruitment materials led participants to believe that an intervention was capable of making them smarter, they may be more engaged in the study, and thus may be more likely to improve than participants who do not expect to benefit.

The use of payment or other forms of extrinsic motivators as a means of incentive may also influence the outcome of an intervention. The sole meta-analysis that examined compensation levels in the context of transfer gain provides some very preliminary evidence of a negative impact of remuneration on transfer improvements, but this effect does not survive the removal of an outlier study (Au et al. 2014). Payment is often used to counter the high dropout rates that accompany cognitive training studies, but few studies consider the implications that high dropout rates might have on a study. One study of cognitive training by Double & Birney (2016) examined the effects of baseline characteristics on both the outcome of training and dropout and found that while individuals who did not complete the training did not differ on motivation-related measures, they

were significantly younger, and less conscientious, than those who completed the study. Furthermore, a study by Jaeggi, Buschkuhl, Shah and Jonides (2013) found that participants who ultimately dropped out of the study also reported being less engaged during training.

Finally, another point related to motivation is the inclusion of “game-like” elements in cognitive training paradigms that are meant to motivate or engage participants. A number of cognitive training programs have been designed to mimic the motivational elements of video games (Jaeggi et al. 2011; Klingberg et al. 2005) while others do not include these game-like features, such as scoring, feedback, or animations. While little work has examined this in the context of cognitive training, there is a considerable body of literature dating back to the 1980s that suggests that these features may have significant impact on one’s experience of a game or program (Malone, 1980). More recently, studies that have attempted to “gamify” education curricula (Aguilar et al., 2014) have found that these elements may improve student engagement, although it is unknown whether these elements would have a similar effect on the outcome of cognitive training. Ultimately, more work is needed to better understand the role of motivation in cognitive training as well as what researchers can do to best motivate their participants. Given that motivation could potentially impact recruitment, drop-out, and performance during the training, these studies provide some impetus for continuing work to investigate participant motivational orientation in the context of cognitive intervention.

### **Research Questions**

This review suggests that none of the individual-difference factors discussed above has been conclusively and consistently found to impact the outcome of cognitive training interventions. All of them are worthy of future study. However, considering the relative lack of direct empirical work that explores motivation and cognitive training, and given the focus and

interest on the relationship between motivation and educational interventions more generally (Schunk, Meece, & Pintrich, 2013), the remainder of this manuscript focuses on three studies that explore the relationship between motivation and the outcome of cognitive training. The first study (Chapter 3) addresses the question of whether training features that are *designed* to be motivational do indeed increase participant motivation and subsequently lead to greater training and transfer performance. The second study (Chapter 4) investigates the preliminary result detailed in Au et al. (2014), that is, does monetary compensation have an undermining effect on the outcome of cognitive training, possibly by negatively impacting one's motivational orientation. The third study (Chapter 5) examines whether a variety of individual-difference factors, including motivation, impact the outcome of cognitive training combined with tDCS stimulation, a technique that has seen increasing use as a means of boosting training effects. A final discussion (Chapter 6) reviews the findings from this body of work and considers the implications for the future of cognitive training research more generally.

## References

- 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 and Review*, 22(2), 366–377.
- Dweck, C. S. (1999). *Self Theories: Their Role in Motivation, Personality, and Development*. Philadelphia: Psychology Press.
- Gruszka, A., Matthews, G., & Szymura, B. (Eds.), *Handbook of individual differences in cognition: Attention, memory, and executive control* (pp. 295–320). New York: Springer
- Hardy, J. L., Nelson, R. A., Thomason, M. E., Sternberg, D. A., Katovich, K., Farzin, F., & Scanlon, M. (2015). Enhancing cognitive abilities with comprehensive training: A large, online, randomized, active-controlled trial. *PLoS One*, 10(9), 1-17.
- Jackson, J. J., Hill, P. L., Payne, B. R., Roberts, B. W., & Stine-Morrow, E. A. (2012). Can an old dog learn (and want to experience) new tricks? Cognitive training increases openness to experience in older adults. *Psychology and aging*, 27(2), 286-292.
- 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 USA*, 105(19), 6829–6833.
- 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 of the USA*, 108(25), 10081-10086.
- Jaeggi, S. M., Buschkuehl, M., Shah, P., & Jonides, J. (2014). The role of individual differences in cognitive training and transfer. *Memory & cognition*, 42(3), 464-480.
- Karbach, J., & Unger, K. (2014). Executive control training from middle childhood to adolescence. *Frontiers in psychology*, 5(390), 1-14.
- Karbach, J., & Verhaeghen, P. (2014). Making working memory work a meta-analysis of executive-control and working memory training in older adults. *Psychological science*, 25(11), 2027-2037.
- Klingberg, T., Fernell, E., Olesen, P. J., Johnson, M., Gustafsson, P., Dahlström, K., Gillberg, C., Forsberg, H., & Westerberg, H. (2005). Computerized training of working memory in children with ADHD—a randomized, controlled trial. *Journal of the American Academy of Child & Adolescent Psychiatry*, 44(2), 177-186.
- Melby-Lervåg, M., & Hulme, C. (2013). Is working memory training effective? A meta-analytic review. *Developmental psychology*, 49(2), 270-291.

- Melby-Lervåg, M., Redick, T. S., & Hulme, C. (2016). Working memory training does not improve performance on measures of intelligence or other measures of “far transfer” evidence from a meta-analytic review. *Perspect Psychol Sci* 11(4), 512–534.
- Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251), 943-951.
- Redick, T. S., Shipstead, Z., Harrison, T. L., Hicks, K. L., Fried, D. E., Hambrick, D. Z., Kane, M. J., & Engle, R. W. (2013). No evidence of intelligence improvement after working memory training: a randomized, placebo-controlled study. *Journal of Experimental Psychology: General*, 142(2), 359-379.
- Richmond, L. L., Morrison, A. B., Chein, J. M., & Olson, I. R. (2011). Working memory training and transfer in older adults. *Psychology and aging*, 26(4), 813-822.
- Salthouse, T.A. (1996). The processing-speed theory of adult age differences in cognition. *Psychological review*, 103(3), 403-428.
- Schaie, K. W., Willis, S. L., & Caskie, G. I. (2004). The Seattle longitudinal study: relationship between personality and cognition. *Aging Neuropsychology and Cognition*, 11(2-3), 304-324.
- Schmiedek, F., Lövdén, M., & Lindenberger, U. (2010). Hundred days of cognitive training enhance broad cognitive abilities in adulthood: findings from the COGITO study. *Frontiers in Aging Neuroscience*, 2(27), 1-10.
- Schwaighofer, M., Fischer, F., & Bühner, M. (2015). Does working memory training transfer? A meta-analysis including training conditions as moderators. *Educational Psychologist*, 50(2), 138-166.
- Shariff, A. F., Willard, A.K., Andersen, T., & Norenzayan, A. (2015). Religious priming a Meta-analysis with a focus on prosociality. *Personality and Social Psychology Review*, 20(1), 27-48.
- Snow, R. E. (1991). Aptitude-treatment interaction as a framework for research on individual differences in psychotherapy. *Journal of Consulting and Clinical Psychology*, 59(2), 205-216.
- Stegenga, J. (2011). Is meta-analysis the platinum standard of evidence? *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, 42(4), 497–507.
- Studer-Luethi, B., Jaeggi, S. M., Buschkuhl, M., & Perrig, W. J. (2012). Influence of neuroticism and conscientiousness on working memory training outcome. *Personality and Individual Differences*, 53(1), 44–49.

- Urbánek, T., & Marček, V. (2015). Investigating the effectiveness of working memory training in the context of Personality Systems Interaction theory. *Psychological Research*.
- Van Elk, M., Matzke, D., Gronau, Q., Guang, M., Vandekerckhove, J., & Wagenmakers, E. J. (2015) Meta-analyses are no substitute for registered replications: a skeptical perspective on religious priming. *Frontiers in Psychology* 6, 1365.
- Weicker, J., Villringer, A., & Thöne-Otto, A. (2016). Can impaired working memory functioning be improved by training? A meta-analysis with a special focus on brain injured patients. *Neuropsychology*, 30(2), 190–212.
- Willis, S.L. (1989). Improvement with cognitive training: Which old dogs learn what tricks? In L. W. Poon, D. C. Rubin, & B. A. Wilson (Eds.), *Everyday cognition in adulthood and late life* (pp. 545-569). New York, NY, US: Cambridge University Press, xii, 708 pp.
- Willis, S. L., & Caskie, G. I. (2013). Reasoning training in the ACTIVE study how much is needed and who benefits? *Journal of aging and health*, 25(8 suppl), 43S-64S.
- Zinke, K., Zeintl, M., Eschen, A., Herzog, C., & Kliegel, M. (2012). Potentials and limits of plasticity induced by working memory training in old-old age. *Gerontology*, 58(1), 79-87.
- Zinke, K., Zeintl, M., Rose, N. S., Putzmann, J., Pydde, A., & Kliegel, M. (2014). Working memory training and transfer in older adults: effects of age, baseline performance, and training gains. *Developmental Psychology*, 50(1), 304-315.

## **CHAPTER 3**

### **Motivational Features in School-based Cognitive Training**

#### **Introduction**

A key challenge in cognitive training research is how to keep participants engaged in training. Training programs are often challenging for participants to complete, and it is expected that they will remain focused on a task or set of tasks for 20 to 40 minutes at a time (Jaeggi et al., 2008; Thompson et al. 2013), for anywhere between a few days (Rueda et al., 2005) to 100 sessions (Schmiedek et al., 2010). Because transfer improvements generally require several hours of training (Jaeggi et al, 2008; Stepankova et al., 2013) it is important that participants in training paradigms remain compliant during training. Additionally, it may be necessary for participants to improve in the training program in order to experience transfer on untrained tests (Jaeggi et al., 2011).

Unfortunately the time commitment and effort required to complete a cognitive training study is often such that many participants do not complete the experiment. Studies often have high dropout rates, including some higher than 25% (Jaeggi et al, 2013; Redick et al., 2013). A variety of individual-difference factors may contribute to a participant's ability to complete the training, such as baseline ability in the training task and one's intrinsic motivation to complete a training program (Jaeggi et al., 2013).

While individual-difference factors are generally outside of the experimenters' control, the design of the training program may also contribute to a participant's engagement in the task,

and these game design elements are often relatively simple to adjust. Cognitive training paradigms vary widely in the type of motivational elements they include, however, and while some studies have focused on recruiting unpaid, intrinsically motivated individuals that may be more likely to engage with and complete a training regimen (Jaeggi et al., 2013), others have utilized substantial financial compensation as a means of encouraging participants to complete the training (Redick et al., 2013; Thompson et al. 2013). Factors that may impact a participant's ability and willingness to engage, comply, and improve in training have been the subject of some interest in recent research. Studies with children often utilize prizes, certificates, and display of high scores to encourage individuals to excel at and complete the training (Holmes, Gathercole, & Dunning., 2009; Jaeggi et al., 2011; Wang et al., under review this issue).

One topic that has not gotten much attention is how game-based motivational elements may contribute to improvements in training and transfer. This is somewhat surprising, considering that elements such as score, tutorials and scaffolding, theming, and feedback are often prominently featured in cognitive training programs. Cognitive psychologists and neuroscientists often find themselves in the role of game designer (Mane & Donchin., 1989; Anguera et al., 2013), and even some of the most basic training paradigms have at least included a motivational chart showing player improvements (Jaeggi et al., 2010). Other training programs, particularly those targeted at children (Jaeggi et al., 2011; Klingberg et al., 2005), look and feel more like traditional video games with appealing art and sound design. Cognitive training games are similar to certain types of entertainment games – specifically, those that Gee (2006) would describe as “problem games,” -- that involve simple, repetitive mechanics, rather than large, open worlds for the player to explore. Almost all tasks used in cognitive training games, from n-back to useful-field-of view to conflict resolution tasks, can be translated into fairly simple

gameplay mechanics (Klingberg et al., 2005; Rueda et al., 2005; Jaeggi et al., 2011; Ball et al., 2010; Alloway, 2012).

While game-based motivational elements have not been well-studied within cognitive training research, some of them have been examined by educational game researchers. Overall, the perspective among these researchers has been that many of these motivational elements aid motivation, and, ultimately, learning (Fishman & Deterding, 2013). For example, one popular game element, persistent scoring (the presentation of a number that represents player performance and changes as the player completes the task successfully) likely encourages engagement and motivation (Toupe, Kerne, & Hamilton 2009). However, the way this scoring is implemented—that is, whether points are earned specifically for completing tasks essential to the learning goal or are awarded for other non-core actions—can determine if scoring hinders or helps learning on the task (Habgood & Ainsworth, 2013). The inclusion of game features may either support or subvert participant motivation to engage depending on how well they tie in with the learning task and the participant's pre-existing motivational framework. For example, imagine a cognitive training game that includes an extra bonus round where players perform some other task non-essential to the training component, such as answering a trivia question. If the number of points possible for the bonus round matches or is greater than that awarded during the core task, participants may be less motivated to perform well during the training portion of the game. This contrasts with situations where the reward is directly reinforcing of the performance task. In one related example, a review of reading incentive programs supported using literacy-related rewards to motivate students (Fawson & Moore, 1999); one study found that students who received a book as a reward following a reading program were more motivated to participate than those who received a token prize (Marinak & Gambrell, 2008).

Psychologists who study motivation are also interested in game-based motivational features (Ryan et al., 2006; Przyblyski, Rigby & Ryan, 2010), possibly because games are an ideal context for understanding the tension between intrinsic motivation and extrinsic reward. Elements such as scoring and feedback may impact a player's intrinsic motivation and may also contribute to his/her success in learning the content included in the game. For example, in Malone's examination of intrinsically and extrinsically motivating game elements, different versions of the game *Breakout* were created that included elements of feedback, such as persistent score and breaking bricks (Malone, 1981). Versions of the game with both of these elements were rated much more highly on a scale of enjoyment by players than versions where they were not present. Theming (referred to as "fantasy" by Malone) was also evaluated and found to significantly contribute to a child's interest in the game, although gender differences were identified in the type of theme each child enjoyed the most.

More recently, some focus has been applied to issues of motivational game elements in cognitive research; however, the research has thus far been inconclusive. Two recent studies compared game versions that included a variety of motivational elements, such as those studied by Malone, to more basic versions of a task. While Prins et al. (2011) found that including game elements such as theming, game-like feedback, and animations increased motivation as well as performance for children completing a working memory training game, recent work from Hawkins et al. (2013), found that the addition of similar game features improved the player experience but not the quality of data collected during a cognitive task. One possible explanation of these mixed results is that the amount of time spent with the game experience also matters – while Prins and colleagues examined the effect of game features over three weekly sessions, the Hawkins et al. study included one single session of play for the games used.

It is also possible that the impact of scoring and other game-like features may differ from game to game, depending on factors such as the goals, difficulty, and demographics of the users. Conclusions drawn from one study cannot necessarily be applied more generally to other types of games or interactive experiences. Nevertheless, no study thus far has systematically examined the impact that *individual* game elements, rather than several features together, have on player performance; previous studies such as those from Hawkins et al (2013) and Prins et al (2011) compare versions of the game with a variety of features to versions of the game without any features present. Therefore, findings from this present research will be of considerable interest to game designers beyond the cognitive training space. By separating out the most popular game-elements included in these training games, such as scoring, lives and leveling, prizes, and theming, we may better understand the extent that these elements contribute to participant engagement and improvements on the task.

To examine how these elements impacted performance on a visuospatial working memory training task, we designed several versions of a three-day working memory game based on a cognitive training task used in previous research (Jaeggi et al., 2011). In this study, the question of interest was whether removing any additional feature might have significant effect on motivation or training gain, and thus each version had one element removed<sup>4</sup>. In the original version of the task, many motivational elements were included, such as changing themes and art, display of score, lives and levels, and prizes and certificates awarded for player compliance and performance. We created new versions of this game, each with one of these elements removed, as well as one with several game-like motivational features absent from the training task. Even

---

<sup>4</sup> However, an alternative design, where a single feature is added to a completely bare-bones version of a task, offers an interesting possibility for future research.

without persistent score, lives, prizes, and changing theme, the task was still *game-like*, with whimsical art and scoring presented between rounds.

This point brings up a significant additional note: why versions of the game with a single element removed were created rather than several versions with one single element added to a bare-bones version of the task. This would likely have been the approach taken if the experimenters had created a completely new game, however, each version of the task is a modification of a training game used in a previous cognitive training study (Jaeggi et al., 2011). Removing a single element generally did not impede gameplay but some elements are interdependent with each other. For example, the prizes students could pick at the end of each day in most conditions were offered based on the total score; students with a higher score could pick prizes of greater value. While other types of feedback (such as the display of leveling on screen) still gave sense of their performance and could be connected to earning better prizes, the addition of performance-based prizes without *any* additional context may not have made sense to the player

We hypothesized that there would be a differential effect of motivational feature for learning on the training task. For example, existing research on the potential negative effects of extrinsic reward, such as Marinak and Gambrell (2008), led us to expect that the removal of prizes might increase learning on the training task. However, in general, the findings from the Prins et al. (2011) study led us to expect that students in the group with all motivational elements included would outperform students in the no motivational element group. Additionally, students in a previous study using the same version of the game as in the “all motivational features” group who reported greater enjoyment of the task outperformed those who did not enjoy the task as much (Jaeggi et al., 2011). The results from Hawkins et al (2013) and Prins et al (2011)

suggested that students in the group with the most motivational elements would rate more highly on self-report measures of intrinsic motivation or enjoyment; it is possible that the versions of the game that students enjoyed more would also be the versions where they experienced greater improvement on the training task. Thus, we expected that removing other features commonly included in games, such as changing theme, scoring, and lives and levels would have a deleterious effect on learning the training task.

We included an outcome measure relatively similar (but not identical) to the training task, in which players were required to identify if a given object presented on screen matched an object presented on screen  $n$ -items earlier. Despite the similarity between the transfer task and the outcome measure, we did not expect to see significant transfer gain due to the limited three-day training duration. Rather, we primarily expected to find differences in player self-reported and observed motivation and performance on the task based on which elements were excluded. We hope that a better understanding of how the game-like elements included in this study impact motivation and performance will help researchers design better, more scientifically useful cognitive training paradigms.

## **Methods**

**Participants.** One hundred twenty-eight students were recruited from seven different school-based summer camps in the southern Michigan region (average age = 10.56 years,  $SD = 2.48$ , range 5 – 14, 37% girls). Students were invited to participate in a three-day experiment in which no compensation was provided outside of the possibility of prizes or certificates in some variants of the intervention; recruitment occurred at tables outside the entrance to the summer camps immediately prior to the start of each camp. Written informed consent was collected from both parents and students prior to participation. Students were also asked if they wished to

continue the experiment prior to each training session and were informed that they could end their participation at any time. Twenty-one students were not included in the analysis due to not completing the entire three days of training and testing ( $N = 13$ ), having taken part in previous cognitive training research ( $N = 2$ ), or being too young to be included in the study (younger than 6 years,  $N = 6$ ). Of the 13 students who dropped out and were not too young or participants in previous cognitive training research, no more than 4 dropped out of any individual condition. One hundred seven students (average age = 10.65 years,  $SD = 2.36$ , range 6 – 14, 44% girls) were then included in the final analysis. Because students completed the tests, questionnaires, and training together as part of the camp, game versions were assigned randomly at the camp level to avoid children comparing the game and prizes amongst themselves and perhaps being disappointed when some received prizes or played more engaging games than others. Running the experiment within summer camps enabled us to evaluate motivational features in a real-world environment, however, one trade-off of this approach is that group sizes and ages differed somewhat depending on which camp students were recruited from. The demographic information for each condition is included in Table 3.1.

Table 3.1. *Demographic information*

<b>Condition</b>	<b><i>N</i></b>	<b>Age (years)</b>	<b>Grade</b>
All optional features included	25	11.28, SD 2.82	6.00, SD 2.38
No theme change	13	10.92, SD 1.71	5.77, SD 1.54
No Points Shown	19	12.21, SD 2.20	6.84, SD 1.64
No Prizes	15	8.40, SD 2.03	3.33, SD 1.63
No Explanation of Lives/Levels	11	9.82, SD 1.72	5.00, SD 1.61
No Explanation of Lives/Levels or Certificates	12	10.83, SD .835	5.42; SD 0.52
No Optional Features Included	12	10.00, SD 1.86	4.83, SD 1.64

**Protocol.** A pre-test was administered on the first day of the experiment prior to the training. The pre-test consisted of a computerized object 2-back assessment that presented participants with a sequence of images one at a time. Participants were required to determine whether each item matched the one presented two items previously and then press one of two keys to indicate their answer. An object was presented every 3 seconds, with a presentation time of 500ms and an inter-stimulus interval of 2,500ms. The pre-test consisted of three blocks of 17 stimuli each and performance was measured as the proportion of correct answers minus the proportion of false responses. Each block included five target trials and ten non-target trials after the presentation of the initial two stimuli. Ten practice trials were included prior to the actual assessment to ensure that the children understood how to complete the task. Following the pre-test on the first day of the study, students began training with the n-back working memory game. After the training on each day, experimenters orally administered brief surveys with Likert-type questions asking how much students enjoyed the game, how exciting the game was, how difficult the game was, and how much effort each student had put into the game. These four questions were adapted for a previous cognitive training study from a factor analysis of the Intrinsic Motivation Inventory (Jaeggi et al., 2011; McAuley, Duncan, & Tammen, 1989). Each of these

variables was averaged over the course of three-days to create enjoyment, excitement, effort, and difficulty variables. Researchers also rated students on how engaged they seemed during each day of training using a Likert-type scale following each training session; this was also averaged over the course of the three days to create a final observer engagement score for each participant. Following the third day of training, participants completed the object 2-back assessment a second time.

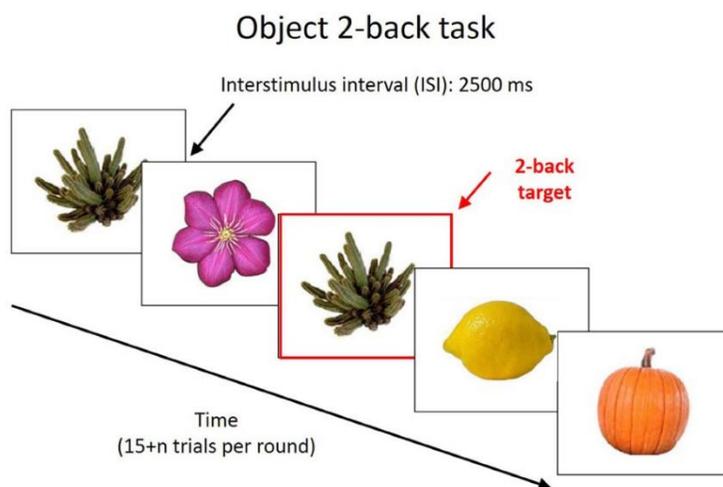


Figure 3.1. Object 2-back task

**Cognitive Training Game.** Participants trained on a game-like computerized working memory task similar to that used in a previous study with children (Jaeggi et al., 2011). This spatial n-back task presented participants with stimuli at one of six locations on the screen, at a rate of 3 seconds each, with 2,500ms between stimuli and with each stimulus presented for 500ms. Students were required to press the A key each time the current stimulus matched the location of the one presented  $n$  items previously, and the L key each time the current stimulus did not match. Participants completed 10 rounds of this task each day, each round consisting of 15 +

$n$  trials, and each round consisting of 5 targets and  $10+n$  non-targets. All versions of the game were adaptive in that the  $n$  level was adjusted depending on performance in each round. If a participant made four or more errors they would lose a single life; after losing three “lives” the participant’s  $n$ -level would be decreased by 1 in the following round. If a participant made three or less errors  $n$  increased by 1 in the following round.

Seven versions of the  $n$ -back training game were developed to examine the role of five motivational features: points, theming, explanation of lives and levels, prizes, and end-of-session certificates. One version of the game included all of these motivation features, while another included none of them. Four of the other versions excluded one of these features. Due to experimenter error, one additional version that was meant to exclude the certificates provided to players at the end of each session also excluded the display of lives and levels feature. However, because this group (with two interrelated elements) was of potential interest, it was included in the subsequent analysis.

**Theming.** Several different themes were developed to make the  $n$ -back task more appealing to students, that is, a frog jumping on lily pads, a cat appearing in windows of a haunted house, and a monkey jumping from sail to sail on a pirate ship (Figure 3.2). In all game versions except for the one excluding theming, the theme changed before the first round on the 2<sup>nd</sup> and 3<sup>rd</sup> day of training. In the “no theme” group as well as the “no motivational features” group, only the lily pad theme was included, and this theme remained persistent across the three days of training.

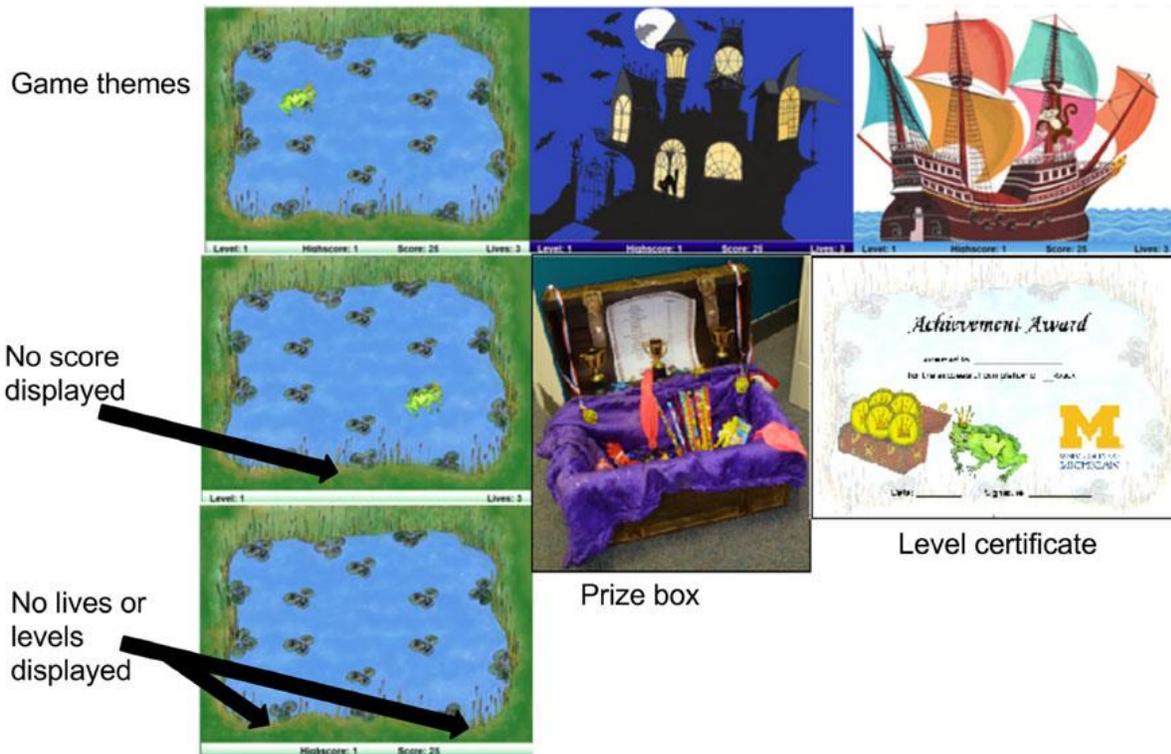


Figure 3.2. Game elements

**Score.** A bar on the bottom of the screen displayed the score as the player completed the n-back task. Points were earned for correctly identifying whether the location of the character on the screen matched the location presented n instances earlier. In versions of the game with prizes, players were instructed that they could trade in points earned for a prize at the end of each day. In the “no points” and “no motivational features” versions of the game, the persistent score was hidden during play (Figure 3.2). The score was still shown at the end of each round, however.

**No display or explanation of lives or levels.** Lives left and the current level were displayed on-screen during play. “Levels” indicated the n level the user was currently on, while “Lives” was used to indicate how many errors the participant could make before dropping an n-back level on the subsequent round. The “Lives” and “Levels” indicator were hidden on the bottom bar (Figure 3.2) for the “no lives or levels explanation” group, as well as the “no

motivational features” group. Additionally, in these groups the experimenter did not mention lives and levels. The game remained adaptive as in the other conditions, however, and the participants still received a certificate after each day’s training summarizing performance.

**Prizes.** Prizes were offered each day after the completion of the game in exchange for “points” the participants had earned. In the “no prizes” group and the “no motivational features” group, participants were given a prize at the very end of the study, but not each day during training. Additionally, participants in those groups were not told that prizes would be given prior to completing the post-test on day three. In the groups where prizes were present, students were allowed to see a treasure box (Figure 3.2) from which they would select items at the end of each day.

**End of session certificates and no display or explanation of lives and levels.** Players were awarded a certificate (Figure 3.2) at the end of each training day celebrating the level they reached. In the “no certificate” version of the game, players were supposed to complete the standard version of the task but without a certificate given at the end of the round, however, the experimenters for this group incorrectly administered a version of the game without the display of lives or levels. Thus players in one of the seven groups were not aware of the role of lives or levels during the task, and additionally did not receive certificates at the end of each day

## Results

**Outlier Analysis.** Outliers in the data were evaluated by examining the average training performance across all 3 sessions for each participant. Using a criterion of 2 SD, we identified two high-performing outliers who trained at an average n-back level of above 4, and two low-performing outliers who performed below at an n-back level of 1.3 or below. However, the

exclusion of these outliers did not impact the outcome of the analyses detailed below, and thus they were not removed from the results described here.

To identify differences in motivation, training performance over time, and pre/post-test performance on the object n-back measure, omnibus ANCOVAs were conducted with all game conditions included; in the case of a significant effect of game-type on these variables, follow-up ANCOVAs were conducted comparing each game variant to the original version with all features included. Despite attempts to recruit summer camps with similar ages, there were significant differences in age across some of the training groups  $F(6,100) = 5.46, p < .001, \eta^2_p = .247$ . Age predicted improvement in the training following a regression analysis including the age of the pooled participants as predictor and the rate of improvement (operationalized as the slope of a linear model – see also below) of the task as outcome,  $\beta = -.202, t(105) = -2.108, p < .05, R^2 = .041$  (proportion of variance in slope explained  $F(1,105) = 4.444, p = .037$ ). Thus, we included age as a covariate in our subsequent analysis.

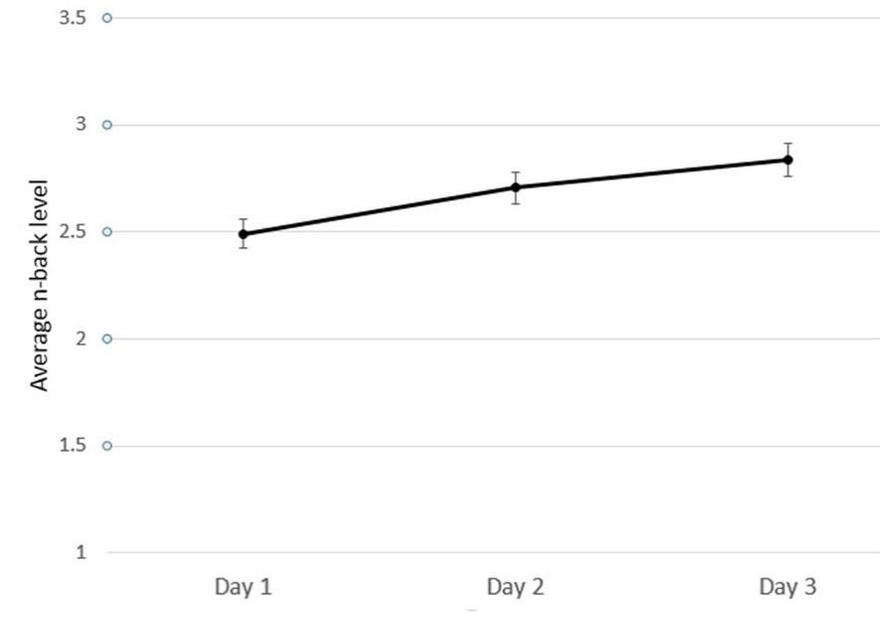


Figure 3.3. Average game performance (n-back level) across all individuals on each day. Error bars represent standard error.

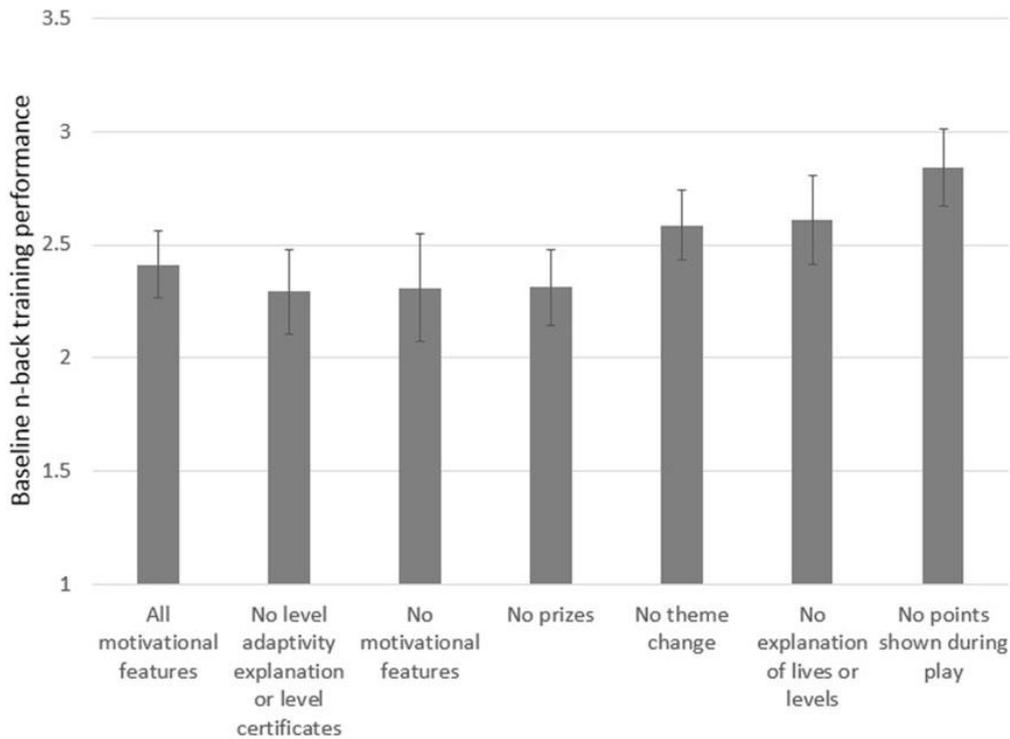


Figure 3.4. Average game performance (n-back level) at baseline for each game version. Error bars represent standard error.

**Training performance.** To quantify each participant’s training improvement over the three sessions of training, the slope of a linear regression model was calculated for each participant using the average n-back level per day of training. Due to the difference in ages across conditions (Table 3.1) and the variance in baseline performance across game versions (Figure 3.4), we included age and starting n-back level as covariates in our analyses. A univariate ANCOVA across conditions revealed a significant effect of game-version on training improvement as measured by linear slope  $F(6,98) = 2.49$ ,  $p = .028$ ,  $\eta^2_p = .132$ ).

To analyze the effect of each individual motivational feature on performance, we then computed a set of univariate ANCOVAs with training slope as the dependent variable, calculated from the average n-back on each day of the training task. We compared students playing the

version of the game with the full set of motivational features to students playing each of the other versions with elements removed, see Table 3.2 and Figure 3.5. Students who played the version of the game without the persistent display of score performed significantly better at the training task over time versus students who played the version of the game with all motivational features,  $F(1,40) = 7.22, p=.010, \eta^2_p = .153$ ) as did students who completed the version of the game without the indication of lives or levels,  $F(1, 32) = 4.48, p=.042, \eta^2_p = .123$ . However, students in the group without theme changes did not perform significantly different from the group with all motivational features  $F(1, 34) = .07, p=.801, \eta^2_p = .002$ ), nor did the group that did not receive prizes after each training session  $F(1, 36) = .01, p=.932, \eta^2_p = .000$ ). The group that did not receive certificates after each day, and also did not see lives or level information during gameplay, trended worse than the all motivational features group, but not significantly so,  $F(1,33) = 2.60, p = .116, \eta^2_p = .073$ . The group that completed the version with no motivational features trended higher but did not differ significantly in performance on the training task from the group with all features,  $F(1, 33) = 2.00, p=.167, \eta^2_p = .057$ ). However, we note that multiple comparisons are a serious limitation here: a Bonferroni correction for these analyses would require a  $p$  value of .0083 or below; none of the results above would survive that correction (although the points versus all motivational features version would come close).

Thus, an additional analysis was performed to further examine the most robust finding, that the display of points on screen may have had a deleterious effect on game performance, as well as to partially address the issue of small samples sizes in the study. A final univariate ANCOVA was thus conducted in a similar manner as above with both the group without any motivational features and the group with only the score removed ( $N = 31$ ) compared to all other participants ( $N = 75$ , all of whom played a version of the game where points were displayed).

This analysis further supported the original finding, as the combined task performance of all individuals who did not see points displayed was significantly better than the combined performance of all individuals who did have points displayed  $F(1,103) = 7.937, p = .006, \eta^2_p = .072$ . A Bayesian follow-up analysis provided moderate confirmation of this primary result with  $BF_{10} = 3.899$  for the model with the points displayed factor and age as a covariate.

Table 3.2. *Training improvement by game variant.*

Condition	<i>N</i>	Estimated mean slope	Standard error	<i>p</i>	partial $\eta^2$
All motivational features	25	0.10	0.07	–	–
No certificates given and no lives or levels displayed	12	–0.08	0.10	n.s.	0.07
No theme changes	13	0.12	0.10	n.s.	0.00
No prizes awarded	15	0.15	0.10	n.s.	0.00
No points displayed	19	0.35	0.08	*	0.15
No lives or levels displayed	11	0.29	0.10	*	0.12
No motivational features	12	0.28	0.10	n.s.	0.06

*Slope here represents the linear model of training improvement calculated using the average n-back performance during each daily session. Slope is controlled for age and baseline n-back level and means are estimated at age = 10.65 and baseline n-back = 2.49. Significance and effect size are drawn from each follow-up ANCOVA that compares a single condition to the original, all motivational features group. Only the “No points displayed” and “No lives or levels displayed” differed significantly from the “all motivational features” comparison group. \* $p < 0.05$ .*

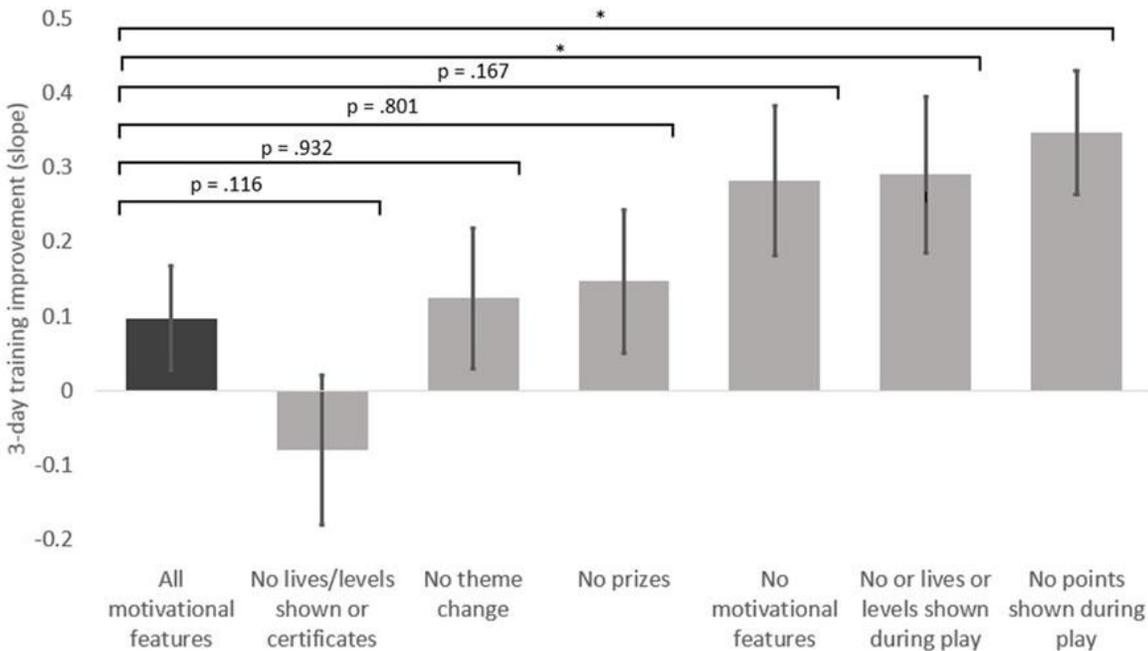


Figure 3.5. Adjusted means of training slope by game-type controlling for age and baseline n-back level. Means are estimated at age = 10.65 and baseline n-back = 2.49. Error bars represent standard error. \*  $p < 0.05$ .

**Motivation.** Participant self-ratings of task-related enjoyment, difficulty, effort and excitement were averaged over the three days and examined as a function of game variant. ANCOVAs with game-type as the independent variable and age as a covariate did not find a significant effect of game-type on student self-report of enjoyment,  $F(6,98) = 1.52$ ,  $p = .180$ ,  $\eta^2_p = .084$ , excitement,  $F(6,98) = 1.43$ ,  $p = .188$ ,  $\eta^2_p = .080$ , or effort,  $F(6,98) = 1.35$ ,  $p = .241$ ,  $\eta^2_p = .076$ , or student self-report of difficulty,  $F(6,98) = 1.94$ ,  $p = .082$ ,  $\eta^2_p = .105$ . However, as in other studies of motivational factors in differently aged students (Lepper, Henderlong-Corpus, & Iyengar, 2005), a median split of students by age revealed that students 10 and under ( $N=47$ , mean = 3.75,  $SD = .79$ ) were significantly more excited to complete the task than students 11 and older ( $N=60$ , mean = 3.29,  $SD = .72$ ) to complete the task ( $F(1,103) = 9.78$ ,  $p = .002$ ,  $\eta^2_p =$

.085). On self-ratings of enjoyment, younger students ( $N=47$ , mean = 3.89,  $SD = .54$ ) were also more likely than older students ( $N=60$ , mean = 3.57,  $SD = .67$ ) to enjoy the task,  $F(1, 103) = 7.38$ ,  $p = .008$ ,  $\eta^2_p = .066$ , suggesting at least that the student questionnaires did accurately capture their personal feelings regarding engagement with the game. Additional analyses of motivational factors for the combined game versions without the display of points compared to the game versions with points on screen did not identify a significant impact of this feature on any of the motivational factors, although students in the group that did not see a persistent score reported applying marginally less effort during gameplay ( $N = 31$ ,  $M = 3.76$ ,  $SD = .55$ ) than those who did see a score ( $N = 75$ ,  $M = 3.48$ ,  $SD = .67$ ),  $F(1,103) = 3.901$ ,  $p = .051$ ,  $\eta^2_p = .036$ . Averaged observer ratings of player engagement over the three-days were also examined as a function of game variant. Again, an ANCOVA was conducted including researcher engagement ratings as the dependent variable, game-type as the independent variable, and age as a covariate. Game-type did not significantly predict experimenter ratings of engagement,  $F(6,99) = 1.91$ ,  $p = .086$ ,  $\eta^2_p = .104$ .

**Object n-back transfer task.** Finally, performance on the object 2-back near-transfer task was examined through an ANCOVA with gain on the object n-back test as the dependent variable, game type as the independent variable, and age and pre-test performance on the object 2-back task as covariates. No differences in improvement were identified between any of the game variants,  $F(6,98) = 1.54$ ,  $p = .175$ ,  $\eta^2_p = .086$ . There was a marginal effect of having score displayed when all individuals who played a version without persistent scoring ( $N = 31$ , mean object n-back gain = .06,  $SD = .21$ ) were compared to the combined participants training with a version with persistent score ( $N = 75$ , mean object n-back gain = .02,  $SD = .29$ ),  $F(1,103) = 3.070$ ,  $p = .083$ ,  $\eta^2_p = .029$ . This is not surprising, however, as the training regimen was likely

too short for sizable near-transfer effects to occur. The untrained object n-back task performance across all participants was not significantly higher after only three days of training ( $M = .473$ ,  $SD = .258$ ) than at the start ( $M = .443$ ,  $SD = .221$ ), as revealed through a paired-samples t-test,  $t(106) = 1.14$ ,  $p = .255$ .

## Discussion

The results of this research should add nuance to our understanding of how popular “motivational” game features impact actual player performance. Over the three days of the study, students playing versions of the game without the persistent display of points and without the display of lives or levels improved significantly more on the game task than students using the original version of the game with all features present. Students playing game versions without changing theme, daily prizes, or end-of-session certificates and the display of lives and levels did not perform significantly differently than the comparison group. Game version did not significantly influence student motivation or performance on the object n-back task.

The effect of these game elements on training performance may seem counterintuitive at first. Why did only the “no score displayed” and “no lives or levels displayed” groups perform differently than the group with all features? And why was the removal of these motivational features associated with improved performance on the training task over the three sessions? It is worth noting that score and lives and levels were indicated on a persistent bar near the game space, a common feature in games. It is quite possible that any element that distracted the user from the challenging n-back task during the actual game would reduce performance, and furthermore, that it might negatively impact learning (i.e., the rate of improvement) over the course of training. This is an interesting finding in light of the fact that cognitive training -- and learning games in general -- often include elements such as score or lives prominently in the

game space. In light of this, one outstanding question is why the group without any motivational features did not perform significantly better than the group with all motivational features included. It is possible that although the no motivational features group did have fewer distracting elements, the exclusion of all other, non-distracting elements had a combined deleterious effect on performance. Determining whether there is a “happy medium” of motivational features that result in optimized performance is a worthwhile goal for future research. Additionally, the other motivational elements, such as awarding prizes and theme changes, did not occur during core gameplay. Over the longer term these elements may impact performance differently, but this finding provides some evidence for removing motivational elements that may be distracting from the player in the early days of a cognitive training regimen.

Overall, the lack of an effect of game variant on student self-ratings of motivation and performance on the untrained object n-back task is not necessarily surprising. Each version of the training program still appeared game-like, and the removal of any individual feature may have a minimal effect on motivation. Alternatively, it may be that certain features that are commonly believed to be “motivational” in nature – such as the display of score – might not actually improve engagement, at least in certain circumstances. Regardless of the underlying reason, this suggests that cognitive game designers may be able to remove some of the game elements that were found to be distracting without any negative impact on a player’s own perceptions of enjoyment and excitement. Finally, it is not necessarily surprising that only the training improvements and not performance on the object n-back task was affected by condition within the limited three-day scope of the study. It is possible that, in an experiment utilizing a much longer training experience, differences in transfer might have been observed.

Several limitations inherent to the present study should be considered. Perhaps of greatest concern is the limited sample size and significant age differences across some of the conditions. While some of the groups are adequately powered, others, due to dropout or other extenuating factors, have as few as 11 participants. Age was included as a covariate in the analyses, but the small sample sizes mean that it is difficult to fully account for the influence of age on differences in training performance. Because these findings would not have survived multiple comparisons, and the effect sizes found were fairly small, these findings must necessarily be seen as preliminary, and, while informative of future research, not conclusive.

Additionally, this is not a true randomized controlled study -- while camps were assigned to conditions randomly, all participants within each camp trained on the same variant of the game. Both of these factors are tradeoffs resulting from the real-world nature of the study; students trained amongst their peers in an actual school environment. Finally, some features of the training regimen, such as the illustrative art style and display of score at the end of each round, exist in all versions of the game. These other features may impact student performance and engagement as well, and were not examined here. The fact that some of the more subtle motivational features, such as persistent score, had a significant impact on three-day performance improvements indicates that these other features should be a focus of future research. As mentioned in the introduction, one further consideration is the possibility that certain game elements may interact with each other and that this may influence participant engagement or performance on the training task. For example, it is possible that persistent scoring is more motivating when participants receive prizes based on their score at the end of each day. This is potentially a significant issue, and one that is not examined in the present study.

Besides including additional game variants, future research could also focus on the impact of these motivational features over a longer-term training regimen. It is possible that some features that impede performance on the training task in this study have less of an effect in a longer training regimen. However, given evidence that long-term improvement in the training task is necessary for transfer gains, any feature that impacts training performance is worth special consideration (Jaeggi et al., 2011). Considering the fact that persistent scoring and the lives/levels feature did impact training performance, we recommend that developers of cognitive training exercise discretion when incorporating these features into their programs.

Our findings have broad implications not only for developers of cognitive training but game designers and cognitive psychologists more generally. Psychologists often make tasks game-like in an effort to drive user engagement. Likewise, within education there has recently been a movement towards game-like formative assessment to evaluate student performance (Wang, 2008). Our findings suggest that game-like elements should be added with caution. Adding game features to an already stressful testing situation may have a deleterious impact on student performance, particularly if the game features add irrelevant cognitive demands. Even seemingly innocuous features, such as displaying score or giving players a certain number of “lives,” may impact performance in a negative fashion. This does not mean that games cannot be effective teaching tools, instruments for cognitive training, or assessment mechanisms. On the contrary, this research provides further support for carefully matching game mechanics and features with the actual task. Researchers might take a look at venerable computerized training task, such as Space Fortress, and examine the impact that non-essential game-like elements included in those tasks have on performance.

While some research has supported the inclusion of game-like elements in cognitive training to improve motivation and training performance (Prins et al., 2011), our findings suggest that these features should be chosen judiciously. Combined with the results from Hawkins et al. (2013), our data suggest that game-like features may not improve the data one collects in research. Furthermore, distracting features may actually impair the participant's ability to improve quickly at the task. Certain "motivational" elements may at best be unnecessary for driving learning on the core task, and at worst have an effect counter to what is intended by the designer. Mae West may have said "the score never interested me, only the game," but persistent display of score, like some other motivational features, might be distracting all the same.

## References

- Alloway, T. (2012). Can interactive working memory training improve learning?. *Journal of Interactive Learning Research*, 23(3), 197-207.
- Anguera, J. A., Boccanfuso, J., Rintoul, J. L., Al-Hashimi, O., Faraji, F., Janowich, J., ... & Gazzaley, A. (2013). Video game training enhances cognitive control in older adults. *Nature*, 501(7465), 97-101.
- Ball, K., Edwards, J. D., Ross, L. A., & McGwin Jr, G. (2010). Cognitive training decreases motor vehicle collision involvement of older drivers. *Journal of the American Geriatrics Society*, 58(11), 2107-2113.
- Fawson, P. C., & Moore, S. A. (1999). Reading incentive programs: Beliefs and practices. *Reading Psychology*, 20(4), 325-340.
- Fishman, B. J., Deterding, S., Vattel, L., Higgen, T., Schenke, K., Sheldon, L., et al (2013). Beyond badges & points: Gameful assessment systems for engagement in formal education. *University of Wisconsin, Wisconsin, Madison*.
- Gee, J. P. (2006). Are video games good for learning? *Nordic Journal of Digital Literacy*, 1(3), 172-182.
- Habgood, M. J., & Ainsworth, S. E. (2011). Motivating children to learn effectively: Exploring the value of intrinsic integration in educational games. *The Journal of the Learning Sciences*, 20(2), 169-206.
- 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-F15.
- Jaeggi, S. M., Buschkuhl, M., Jonides, J., & Shah, P. (2011). Short-and long-term benefits of cognitive training. *Proceedings of the National Academy of Sciences of the USA*, 108(25), 10081-10086.
- Jaeggi, S. M., Buschkuhl, M., Jonides, J., & Perrig, W. J. (2008). Improving fluid intelligence with training on working memory. *Proceedings of the National Academy of Sciences of the USA*, 105(19), 6829-6833.
- Jaeggi, S. M., Buschkuhl, M., Shah, P., & Jonides, J. (2013). The role of individual differences in cognitive training and transfer. *Memory & Cognition*, 1-17.
- 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 & Adolescent Psychiatry*, 44(2), 177-186.

- Mané, A., & Donchin, E. (1989). The space fortress game. *Acta psychologica*, 71(1), 17-22.
- Malone, T. W. (1981). Toward a theory of intrinsically motivating instruction\*. *Cognitive science*, 5(4), 333-369.
- Marinak, B. A., & Gambrell, L. B. (2008). Intrinsic motivation and rewards: What sustains young children's engagement with text?. *Literacy Research and Instruction*, 47(1), 9-26.
- Prins, P. J., Doyis, S., Ponsioen, A., Ten Brink, E., & Van der Oord, S. (2011). Does computerized working memory training with game elements enhance motivation and training efficacy in children with ADHD?. *Cyberpsychology, Behavior, and Social Networking*, 14(3), 115-122.
- Przybylski, A. K., Rigby, C. S., & Ryan, R. M. (2010). A motivational model of video game engagement. *Review of General Psychology*, 14(2), 154.
- Redick, T. S., Shipstead, Z., Harrison, T. L., Hicks, K. L., Fried, D. E., Hambrick, D. Z., ... & Engle, R. W. (2013). No evidence of intelligence improvement after working memory training: A randomized, placebo-controlled study. *Journal of Experimental Psychology: General*, 142(2), 359.
- Rueda, M. R., Rothbart, M. K., McCandliss, B. D., Saccomanno, L., & Posner, M. I. (2005). Training, maturation, and genetic influences on the development of executive attention. *Proceedings of the national Academy of Sciences of the USA*, 102(41), 14931-14936.
- Ryan, R. M., Rigby, C. S., & Przybylski, A. (2006). The motivational pull of video games: A self-determination theory approach. *Motivation and Emotion*, 30(4), 344-360.
- Thompson, T. W., Waskom, M. L., Garel, K. L. A., Cardenas-Iniguez, C., Reynolds, G. O., Winter, R., ... & Gabrieli, J. D. (2013). Failure of working memory training to enhance cognition or intelligence. *PloS One*, 8(5), e63614.
- Schmiedek, F., Lövdén, M., & Lindenberger, U. (2010). Hundred days of cognitive training enhance broad cognitive abilities in adulthood: findings from the COGITO study. *Frontiers in Aging Neuroscience*, 2.
- Stepankova, H., Lukavsky, J., Buschkuehl, M., Kopecek, M., Ripova, D., & Jaeggi, S. M. (2013). The malleability of working memory and visuospatial skills: A randomized controlled study in older adults.
- Toups, Z. O., Kerne, A., & Hamilton, W. (2009). Motivating play through score. In *ACM Computer Human Interaction 2009 Workshop on Engagement by Design* (pp. 4-9).

## **CHAPTER 4**

### **Cognitive Training and Monetary Compensation**

#### **Introduction**

The goal of cognitive training is to improve a single cognitive skill, or a set of skills, that would then increase performance across a wide range of untrained tasks that draw on those skills. As discussed earlier, however, outcomes from cognitive training studies are often mixed. It is common for only some individuals to benefit from the intervention, and it is largely unknown what individual-difference factors play a role in the success of these interventions. Thus, the determination of which factors moderate task-based improvements and transfer following cognitive training, and by how much, is an important question. While a wide variety of factors, such as the degree of researcher involvement, subject setting, training duration, and feedback to participants have been proposed as potential moderators of improvements (Jaeggi et al., 2014), motivational orientation has been shown to be an important factor in determining the success of cognitive interventions and learning more generally (Pintrich, 1999). This has only recently been explored in the context of cognitive training research (Prins, 2011), however. The goal of the present study is to investigate the impact of monetary compensation and motivation on the outcome of one widely used cognitive training intervention: working memory (WM) training.

One potential explanation for differences in outcomes across studies is monetary compensation, which may impact one's motivation to complete a task (Murayama et al., 2010).

Many WM training studies have utilized the dual n-back task, which requires participants to judge whether items presented in an auditory stream are the same as those presented n-items previously and simultaneously judge whether locations presented in a spatial stream are the same as those presented n-items previously. As participants succeed at a particular n-level, n is incrementally increased. Training on this difficult, adaptive WM task was found in one prominent study to lead to improvements in a measure of fluid intelligence (Gf; Jaeggi et al., 2008). This and other similar n-back training paradigms have subsequently been used by several research teams, and while some studies have also found transfer to Gf or related visuospatial reasoning measures (Colom et al., 2013; Jaeggi et al., 2011; Jaeggi et al., 2014; Jausovec & Jausovec, 2012; Rudebeck et al., 2012; Schweizer et al., 2011; Stephenson & Halpern, 2013), other studies have not replicated these transfer effects (Chooi & Thompson, 2012; Redick et al., 2013; Thompson et al., 2013).

We propose that for working memory training, as in a wide range of learning tasks, the degree to which individuals are intrinsically motivated will have a substantial impact on learning, performance, and achievement (Benware & Deci, 1984; Condry & Chambers, 1978; Lin et al., 2001; Spence & Helmreich, 1983). Participants who lack intrinsic motivation are less attentive and more distractible (Fransson, 1977), do not maximize effort (Vollmeyer & Rheinberg, 2000), and are more likely to disengage when a task becomes difficult (Dev, 1997). There is evidence that extrinsic reward may have a deleterious effect on compliance in completing repetitive cognitive tasks (Murayama et al., 2010) as well as reducing performance on attentional measures (Robinson et al., 2012). Additionally, the existence of extrinsic motivators, such as payment for participation, may undermine intrinsic motivation in a wide range of contexts (Blumenfeld, Kempler, & Krajcik, 2006; Henderlong & Lepper, 2002; Pittman & Heller, 1987; Tang & Hall,

1995). However, evidence in support of the hypothesis that extrinsic rewards undermine performance is mixed. While a meta-analysis by Deci, Koestner, & Ryan (1999) did identify a negative impact to self-determination and therefore intrinsic motivation, another meta-analysis by Cameron & Pierce (1994) found inconsistent effects of extrinsic reward on intrinsic motivation. Furthermore, individual differences may moderate the degree to which extrinsic factors may be undermining. For example, Robinson et al. (2012) suggest that individuals who express very high intrinsic motivation may be less susceptible to the negative effects of providing an extrinsic reward. Finally, not all extrinsic rewards are alike: nominal extrinsic rewards, such as a small trinket or cash prize of only a few dollars, may not negatively impact intrinsic motivation (Ross, 1975). Additionally, only tangible rewards, such as money or physical prizes, have been shown to consistently undermine intrinsic motivation (Deci, 1971, 1972; Reeve & Deci, 1996; Swann & Pittman, 1977).

Given that the impact of extrinsic rewards is complex and dependent on a variety of factors, including individual differences, the type of extrinsic reward, and the type of task, it may be valuable to consider the influence of extrinsic rewards in the context of WM training and its efficacy on transfer. In other fields, such as exercise compliance, some research suggests that receiving extrinsic rewards may reduce one's intrinsic motivation to adhere to a training regimen (Frederick & Ryan, 1995). Because WM training studies are often time-intensive and challenging for participants, researchers frequently provide substantial payment. More than \$100 has been provided to participants in many of these studies as a means of recruitment (Anguera et al., 2012; Kundu et al., 2013; Redick et al., 2013; Rudebeck et al., 2012; Schweizer et al., 2011; Thompson et al., 2013). Other studies have provided nominal or no payment to subjects for their participation (Jaeggi et al., 2008; Jaeggi et al., 2010; Jaeggi et al., 2014; Jausovec & Jausovec,

2012; Stepankova et al., in press; Stephenson & Halpern, 2013). Compensation in these studies is often contingent simply on compliance or on a per-session basis (Redick et al., 2013); as discussed in the review of the motivation literature above, some have suggested that this compensation method may negatively impact intrinsic motivation, and thus performance on the tasks and assessments (Deci, Koestner, & Ryan, 1999). One recent meta-analysis of dual n-back training research provides some evidence that compensation may be negatively correlated with transfer gains, although this effect did not survive the removal of outliers (Au et al., 2014). Of course, these WM training studies also differ on other dimensions as well, and it remains difficult to determine why some of these studies demonstrated transfer following cognitive training and others did not. Nevertheless, it is conceivable that the amount of payment might be an important factor that contributes to the magnitude of transfer. Thus, in the current study we directly manipulate the effect of extrinsic motivation on training and transfer outcomes through the use of compensated and uncompensated groups.

As discussed earlier, there are likely individual differences in the extent to which extrinsic motivators might reduce intrinsic motivation. Consider an individual who is highly intrinsically motivated by a particular task (e.g. an academic who enjoys conducting his or her research). Compensation for those individuals may not necessarily undermine intrinsic motivation. Amabile et al. (1994) and Durik & Harackiewicz (2007) suggest that individual differences in motivational orientation may determine whether or not extrinsic reward has an undermining effect. Beyond the effects on task compliance and intrinsic motivation itself, some research on motivation has found that under some conditions, such as during an experimental session rather than in a participant's free-time, extrinsic motivation may not have a detrimental effect on participant performance, and may even be beneficial (Wiersma, 1992). We predicted

that participants who report being intrinsically motivated despite being compensated should perform as well as those who participated on the basis of intrinsic motivation alone. We directly asked participants in the compensated group why they participated in the study as a final question following post-testing, and categorized their responses based on whether or not they mentioned improving cognition as a reason for joining the study. This stated interest in improving cognition may serve as a proxy for intrinsic motivation.

Personal beliefs and personality factors, such as neuroticism and conscientiousness, have also been found to be linked to motivation as well as training performance (Judge & Ilies, 2002; Studer-Luethi et al., 2012). There is some evidence that transfer improvements following cognitive training may be a result of placebo effects rather than a function of some core cognitive ability being increased. In a study by Foroughi and colleagues (2016), participants were recruited by responding to one of two flyers: one that emphasized improvements to IQ, and another that focused on receiving subject pool credits. Participants in the group that viewed IQ-focused flyers had higher scores on the Theories of Intelligence scale at baseline, and also performed better than the non-IQ recruited group on two matrices-based tests following a single training session. Although this study included only a single session of training, another study by Jaeggi et al. (2014) found a similar association between incremental theories of intelligence and the amount of improvement on a visuospatial reasoning factor following training. This effect was driven principally by the active control group, suggesting the existence of a placebo effect. In that study, the impact of training on transfer remained significant after controlling for beliefs about intelligence; however, it is possible that self-theories about cognition and development may significantly impact performance, at least in some instances (Dweck & Master, 2008). Because we have included the same questionnaires assessing motivation as in Jaeggi et al. (2014), we can

assess how these factors might interact with the effects of monetary compensation. For example, trait factors such as those measured by the Cognitive Failures Questionnaire may impact a subject's motivation to participate, as individuals with greater numbers of reported cognitive failures may be self-motivated to improve through the training. Conditions such as this may play a role in determining whether extrinsic motivation negatively affects training performance and transfer.

Thus the primary goal of the present study was to test the effect of extrinsic motivation on training outcomes. Thus we recruited individuals to participate in a well-paid study (\$352) for approximately 12 hours of combined testing and cognitive training. Neither the recruitment flyer nor initial screening for this group included any "brain training" or cognitive improvement messaging. This amount of compensation is identical to that used in Redick et al.'s (2013) study, a non-replication of Jaeggi et al.'s original dual n-back research (2008). We compared the effectiveness of training and transfer in this new compensated group to a group of uncompensated individuals recruited for another recent study of cognitive training and individual differences (Jaeggi et al., 2014) that was identical across all methodological features (training task, transfer tasks, experimental conditions) except for compensation and recruitment. In Jaeggi et al. (2014), participants were recruited to participate in a study of brain training and there was no compensation provided. We predicted that participants who were recruited to the current paid study would be less intrinsically motivated to engage in training, and consequentially, they would show reduced training and transfer effects compared to those who were recruited for a "brain training" study and thus were intrinsically motivated to train and improve cognitive function.

## Methods

**Participants.** Forty-six participants were recruited for the current study from the Ann Arbor, Michigan community (mean age: 21.39 years; SD: 3.13; range: 18-30; 24 women). The participants were recruited to participate in a paid study (as opposed to a “cognitive training” or “brain training” study) to complete a set of computerized tasks and were compensated \$352 to complete the study. Two participants were excluded before finishing the pre-test due to technical problems, and one participant was asked to leave the study following two sessions of training after remembering that he had participated in cognitive training research previously (participants were screened during recruitment to ensure they had not participated in WM training research before). Of the remaining 43 individuals, 1 participant withdrew from the study after completing no more than 3 sessions of training. Another 6 participants dropped out at some point during the training and/or failed to complete the post-test, after having trained for 10.83 sessions on average (SD: 6.27; range 4-20). The final group of participants who completed pre- and post-testing and a minimum of 17 training sessions consisted of 36 individuals (mean age: 21.03 years; SD: 3.05; range: 18-30; 21 women).

We compared training and transfer performance of this paid group of participants to data originally published in Jaeggi et al. (2014). The Jaeggi et al. participants were recruited from the Ann Arbor community to participate in an experiment advertised as a “Brain Training Study” without the offer of compensation or class credit. Included in the current analysis are two of the three conditions from that study: an unpaid dual n-back group and an unpaid active control group. Specifically, we included data from the 52 participants in those two conditions who completed at least 17 sessions of training (mean age: 24.58 years; SD: 5.72; range: 18-43; 28 women; see Table 4.1 for detailed information on recruitment and dropouts). Three participants

from the earlier groups failed to complete a single questionnaire each and two participants failed to complete a single assessment each, these participants were included in all portions of the analysis except those involving those assessments or questionnaires. There were no significant demographic differences between the final sample of paid and unpaid participants except that unpaid participants were significantly older than paid participants,  $F(1,86) = 11.559, p = .001, \eta^2 = .118$ ). It is possible that the compensation may have contributed to this age difference; however, a linear regression analysis revealed that age was unrelated to training or transfer outcomes. Sample sizes in this study are slightly larger but comparable to the groups used in Redick et al. (2013), both for the purposes of comparison and achieving adequate statistical power.

Table 4.1. Attrition rates for all individuals who completed at least 3 sessions of training.

	Completed Training		Dropped Out	
	N	Sessions	N	Sessions
<b>Dual N-back: unpaid</b> (Percentage) (SD; range)	25 58%	19.6 1.08; 17-22	18 42%	13.28 5.45; 5-20
<b>Dual N-back: paid</b> (Percentage) (SD; range)	36 86%	20 0; 20-20	6 14%	10.83 6.27; 4-20
<b>Knowledge Trainer</b> (Percentage) (SD; range)	27 75%	19.89 .58; 18-21	9 25%	10.78 7.76; 4-20

**Recruitment.** To recruit participants for this study, we posted flyers throughout the Ann Arbor area. The flyers used to recruit participants for this study were different from those used to recruit participants in the Jaeggi et al. (2014) study. Both flyers are included in Figure 4.1. The paid cohort was offered compensation in a manner comparable to previous paid training studies, particularly Redick et al. (2013). Participants were offered \$72 for completing the 3-hour pre-test

session, \$10 for each training session completed, and an additional \$72 for completing the 3-hour post-test session. Participants were paid following the post-test or after dropping out of the study. The paid condition data was collected one academic year after the data from Jaeggi et al. (2014) were collected. However, to mitigate the potential cohort effects between the different conditions, flyers recruiting participants were placed in identical locations to Jaeggi et al. (2014) and all testing materials and protocols were kept identical. Besides the recruitment flyer and payment, the only difference in protocol between the two studies was a final manipulation-check question administered to the paid condition after all other post-testing was completed. In both studies, the only mention of improving cognition between recruitment and the debriefing was a single sentence in the consent form: “The ultimate goal of this research study is to better understand how we can improve cognitive functions by means of a computerized training intervention.”

### Participate in a Computerized Study Earn Up to \$352



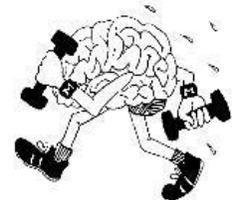
We are currently looking for research subjects to participate in a study involving computerized tasks.

The study requires participants to complete 15-25 minutes of computerized tasks per day over a period of four weeks. The tasks will be completed in the convenience of your home. Before and after completing the tasks, different cognitive functions will be assessed.

Monetary compensation of up to \$352 will be provided for participants who complete all tasks and assessment sessions.

If you are between 18-45 years old and are interested in participating, or if you would like more information, please contact our lab at [tasksforcash@umich.edu](mailto:tasksforcash@umich.edu)

### Train Your Brain and Get Smarter!



Want an opportunity to improve your memory and intelligence while tracking your success with a **scientifically proven** training program?

We are currently looking for interested and motivated people who want to increase their mental capacity **FOR FREE!!**

Our research has been featured in the **New York Times** and was named **Discover** magazine's top 100 stories of 2008!

The study requires 15-25 minutes of training per day over a period of four weeks. The training is computerized and will take place at the convenience of your home. Before and after the training, different cognitive functions will be assessed in order to determine the effectiveness of the training.

If you are between 18-45 years old and interested in training your brain, or if you would like more information, please contact our lab at [trainyourbrain@umich.edu](mailto:trainyourbrain@umich.edu).

Figure 4.1. Recruitment posters used for the paid and unpaid groups.

**Procedure.** Participants in the paid dual n-back group underwent the identical procedure and completed the same battery of tasks as in the Jaeggi et al. (2014) study. Following written informed consent, participants completed an initial baseline assessment session, consisting of five reasoning tasks (see below) administered in the following order: Inferences, Surface Development, Verbal Analogies, APM, Culture Fair Intelligence Test (CFT). This testing session lasted approximately 90 minutes. Due to the length and intensity of these testing sessions, participants were allowed to take breaks between successive assessments if they wished. In the second session, the remaining assessments were administered in the following order: Space Relations, Reading Comprehension, Form Board, Digit Symbol, Bochumer Matrizen-Test (BOMAT). This second pre-test session lasted approximately 90 minutes as well. Participants were also asked to complete questionnaires administered via computer between the first and second session (Need for Cognition (NFC), Theories of Cognitive Abilities (TOCA), Cognitive Failures Questionnaire – Memory and Attention Lapses (CFQ), Personality Questionnaire). After having completed all assessments described above, participants in the paid dual n-back group (and unpaid dual n-back group in Jaeggi et al. (2014)) were instructed on the dual n-back training task. Similarly, participants in the knowledge-training group (active control group) received training on their training task. All participants were instructed to train once a day and at least five times a week for a total of 20 sessions; in fact, all participants in the paid dual n-back group completed the full 20 sessions (this was not the case in the unpaid study). In order to increase and monitor compliance, participants were asked to email their training data files to the experimenters each day. Reminder emails were sent if participants failed to do so. After completing up to 20 sessions of training, participants completed two post-test sessions in a similar manner as for the pre-test, with the notable exception that they worked through the

alternate version of the assessment for counterbalancing purposes. A/B versions of tests were counterbalanced across participants, and parallel-test versions were used in order to avoid participants encountering the same items twice.

Following completion of the post-test, participants in the paid dual n-back group were asked an additional manipulation-check question regarding their reasons for signing up for the experiment. These responses were coded for whether or not they included a mention of extrinsic motivation only (compensation) or whether they also included some mention of intrinsic motivation, such as improving cognition. Inter-rater reliability across three researchers was high (Krippendorff's alpha reliability = .82, N decisions = 108). Twenty-one (58%) participants mentioned only extrinsic motivation, while 15 (42%) participants mentioned some intrinsic motivation for joining the experiment. We note that the participants who mentioned intrinsic motivation did so despite the sign-up and recruitment for the study including no mention of personal improvement. This suggests that these participants developed this intrinsic motivation after deciding to join the study.

**Transfer Measures.** *AFOQT Reading Comprehension Test* (Air Force Officer Qualifying Test; Berger, Gupta, Berger & Skinner, 1990 as used in; Kane et al., 2004; Kane & Miyake, 2007). Participants completed ten items per test version within a five-minute time limit. Each item requires participants to read a short paragraph and then choose one of five possible answers that will complete the final sentence of the paragraph. Score was total items correct.

*ETS Inferences Test* (Ekstrom et al., 1976; cf. Kane et al., 2004). Participants were given six minutes to answer ten items (either the odd or even items from the original ETS assessment). Each individual item consists of one to two short sentences and five possible conclusions.

Participants must determine which conclusion can be determined from the statement without needing additional information.

*Verbal Analogies Test* (based on Kane et al., 2004; Wright, Thompson, Ganis, Newcombe, & Kosslyn, 2008). This computerized assessment requires participants to view two sets of word pairs and determine whether a word pair presented on the left side of the screen is the same (by pressing the 1 key) or different (by pressing the 0 key) from the word pair presented on the right side. The relationship between words in a pair consisted of a synonym, category, function, opposite, or linear order. Participants must respond to each trial within 8 seconds, although the task itself was self-paced. A 500ms black screen was presented between trials. Each participant completed a total of 57 distinct trials after 8 practice trials. Score was recorded as the proportion of items solved correctly.

*Bochumer Matrizen Test* (BOMAT; Hossiep, Turck, & Hasella, 1999). The BOMAT consists of a set of visual reasoning puzzles ordered in increasing difficulty and developed to be challenging for high-ability participants. Each puzzle is a 5x3 matrix with a single missing section. Participants must review 6 possible answers to complete the matrix and choose the best option. Participants first complete 10 items for practice and then 27 items on the actual test drawn from the long (80 item) version of the assessment. The score was the number of correct items minus the first of the 27 test items; this initial item was considered additional practice. No time restriction was used in this task.

*Cattell's Culture Fair Test* (CFT; Cattell & Cattell, 1963). The present experiment utilized Scale 3 Forms A and B, which included 22 practice items plus 100 test items total. Each version consists of four subtests. The tasks on the subtests include series, classification, matrices, and conditions (topology) (cf. Johnson, Nijenhuis, & Bouchard, 2008). Using the combined

items from Forms A and B, three versions of the assessment with an equal number of items from each subtest (8-10) were created. Participants completed 2-3 practice items for each subtest and then solved the remaining 34 items without time constraints. Score was recorded as the total number of correct answers.

*DAT Space Relations* (Bennett, Seashore, & Wesman, 1972; cf. Kane et al., 2004). Each item of the DAT consists of the outline of a pattern that can be folded into a 3-D object. Participants then review four possible answers and choose the one object that can be folded from the outline shown. The original test was divided into odd (version A) and even (version B) items so that two versions were created with 17 items each (the last item from the original test was left out). A 5-minute time limit was used and the score consisted of the total number of items completed correctly during this time.

*ETS Form Board Test* (Ekstrom et al., 1976; cf. Kane et al., 2004). For each item participants review a group of five 2-D shapes. Some or all of these shapes, when combined, create a 2-D shape shown at the top of the page. Participants mark a plus sign under the shapes that can be used to create the shape at the top of the page and a minus sign for those that should be left out. Each target shape includes six sets of five items below it, and each version of the test included four different figures: cross, square, triangle, pentagon. Version A of the assessment included items 1-6, 19-24, and 31-42 from the original test, while version B included items 7-18, 25-30, and 43-49. Participants first complete 2 practice items and then are given 8 minutes to solve 24 test items, each consisting of 5 possible shapes, for a total of 120 answers. Total score was the number of correct answers given in the allotted time.

*ETS Surface Development Test* (Ekstrom et al., 1976; cf. Kane et al., 2004). Participants are given a set of 2-D drawings that can be folded along dotted lines to create a 3-D shape. Some

of the edges of each drawing are marked with letters, and some with numbers. For each item participants must determine which of the edges with letters will match the edges with numbers. Each item includes five numbered edges, and each assessment includes six items, for a total of 30 responses. After completing a practice item participants must complete the assessment in 6 minutes. The odd items of the original test were used for version A, while version B was drawn from the even items. Score was the total number of correct items given during the time allotted.

*Raven's Advanced Progressive Matrices (APM; Raven, 1990).* Like the BOMAT, this test is composed of a set of visual reasoning problems ordered by level of challenge. Each item consists of a 3x3 matrix in which the tile in the lower right corner is missing. Participants must review 8 possible responses and choose the tile that successfully completes the empty slot. Set 1 from the Raven's assessment was completed as practice, and half the items on Set 2 were used for each of two versions of this test. No time limit was used for this assessment; however, the experimenters note that this factor, combined with the low number of items on the test, may cause potential ceiling effects and make measuring transfer more difficult. The score was the total number of correct responses.

*Digit Symbol Test.* The Digit Symbol test from the WAIS (Wechsler, 1997) consists of the presentation of nine digit-symbol pairs. Participants must quickly examine each digit and then fill in the corresponding symbol under a list of 130 digits. Participants have 90 seconds to complete as many items as possible, and this total is used as the score.

**Questionnaires.** *Need for Cognition (NFC; Cacioppo & Petty, 1982).* This survey includes statements such as "I really enjoy a task that involves coming up with new solutions to problems." Participants were asked to indicate their level of agreement or disagreement on a 9-

point Likert-type scale; the survey was then used to determine how much a subject enjoys cognitively difficult activities.

*Theories of Cognitive Abilities* (TOCA; Dweck, 1999). We measured the extent to which participants think of intelligence as a fixed or changeable trait. The questionnaire consists of 8 statements, such as “You have a certain amount of cognitive ability and you can’t really do much to change it.” Participants indicate their agreement or disagreement on a 6-point Likert-type scale.

*Cognitive Failure Questionnaire – Memory and Attention Lapses* (CFQ-MAL as used in McVay & Kane, 2009). Participants’ proclivity for making cognitive errors was measured using a list of 40 questions, such as “Do you read something and find you haven’t been thinking about it, so you have to read it again?” Responses are given on a 5-point Likert-type scale.

*Personality Inventory*. Personality traits were measured using the Mini-Marker Set based on the Five-Factor Personality Model (Saucier, 1994). On this test, subjects indicated their level of agreement with 40 single adjective descriptors of personality on a 5-point Likert scale. In a previous study of n-back working memory training, neuroticism and conscientiousness both impacted the outcomes of the training; for example, participants training on the dual n-back task with low neuroticism displayed greater transfer than those with high neuroticism (Studer-Luethi, et al., 2012).

*Post Test Questionnaire*. Following the post-test, participants were asked questions such as whether they enjoyed the training and if they believed the intervention improved their memory or attention. In the paid dual n-back group, an additional final question was included at the end of the post-test questionnaire regarding their motivation for joining the study.

**Training Tasks.** *Dual N-back Task.* The dual n-back working memory training task was identical to the one used in Jaeggi et al. (2014). In this task, participants processed two streams of stimuli (auditory and visuospatial; 8 stimuli per modality) that were synchronously presented at the rate of 3s per stimulus. Participants must continuously determine whether each successive item (auditory and visual) matches the item that came 1, 2, 3, or n-back before it. After each block of trials, the n level was varied as a function of performance. If participants made fewer than 3 errors in both modalities, the level of n increased in the next round by one, if they made more than 5 errors in either modality, the level of n decreased in the next round by one, and in all other cases, n remained the same. There were 6 targets per modality in each round. Participants trained for 15 rounds in each session, each round consisting of  $20 + n$  trials. We examined the nature of the training function from two dual n-back cohorts used in this study and found that a logarithmic model, rather than a linear or exponential model, provided the best fit. Therefore a logarithmic model was used to calculate each individual's training performance over the course of the intervention.

*Knowledge-Training Task.* The knowledge training control task was identical to that used in Jaeggi et al. (2014). Participants completed GRE-style general knowledge, vocabulary, and trivia questions selected from a pool of approximately 5,000 questions. Each question was displayed in the center of the screen, and participants chose one of four answer alternatives presented below the question. After the participant's response, the correct answer was provided, along with some additional facts related to the question. Incorrectly answered items were presented again in the beginning of the next session in order to evoke a learning experience. Users generally get many questions incorrect in the initial sessions of the task before consistently entering more correct answers in subsequent days of training. This sort of presentation is

responsible for a training curve that does not fit the logarithmic model used for the dual n-back task.

A training session in either condition lasted approximately 20-30 minutes. After each session, participants rated how engaged they were during the session (responses were given on a Likert-type scale ranging from 1-9). At the end of each session participants were presented with a curve displaying their average N (one point for each session) in relation to a pre-set curve that was derived from previous data collected in our laboratory.

Jaeggi et al. (2014) performed an exploratory factor analysis with an oblique rotation, utilizing pre-test measures from participants in their study and identified two factors that served as the basis of the current transfer analyses. The first factor, determined to be a verbal reasoning component, included four measures that accounted for 35% of the variance. These measures included the CFT, ETS Inferences, AFOQT Reading Comprehension, and Verbal Analogies. The second factor, inclusive of five measures, accounted for 13% of the variance. These measures included the APM, BOMAT, Form Board, Space Relations, and the Surface Development test. Because the Digit Symbol Test did not load on either of the factors, it was not included in the subsequent analysis. Composite scores (the mean of the standardized gains) were calculated for each factor. In the present study, because we had a small sample size, we utilized the same factor structure as Jaeggi et al. (2014). Pre- and post-test scores for each measure were divided by the standard deviation of the whole sample at pre-test for each measure and then averaged into their respective factors (cf. Jaeggi et al., 2014). For the transfer measures the current analysis utilized repeated-measures ANCOVAs with test version as a covariate to evaluate changes in performance on the untrained cognitive assessments following the intervention.

## Results and Discussion

**Outlier Analysis.** Outliers in the data were evaluated by examining the average training performance for each participant. Using a criterion of 2 SD, we identified two low-performing outliers who performed below at an n-back level of 1.4 and 1.25. However, the exclusion of these outliers did not impact the outcome of the analyses detailed below, and thus they were not removed for the analyses described here.

**Effect of Payment on Traits.** A multivariate analysis of variance (MANOVA) examining the paid (dual n-back) versus unpaid (dual n-back and knowledge trainer) groups and eight trait measures (extraversion, agreeableness, conscientiousness, emotional stability, intellectual openness, cognitive failures, TOCA, and Need for Cognition, see Table 4.2) revealed significant group differences at pre-test on personality and motivational factors for participants who completed the training,  $F(8,77) = 4.508, p < .001, \text{Wilks' } \Lambda = .681; \eta^2_p = .319$ . Univariate ANOVAs were then conducted for each measure using a Bonferroni adjusted  $\alpha$  level of .00625 (.05/8). For the personality factors, there was no difference for Extraversion rating  $F(1,84) = 2.635, p = .108, \eta^2_p = .030$ , Intellectual Openness  $F(1,84) = .704, p = .404, \eta^2_p = .008$ , Agreeableness  $F(1,84) = 1.049, p = .309, \eta^2_p = .012$ , or Emotional stability  $F(1,84) = 6.885, p = .010, \eta^2_p = .076$ . Additionally, there was no difference between the paid and unpaid groups on the Theories of Cognitive Abilities measure  $F(1,84) = 2.389, p = .126, \eta^2_p = .028$  or Need for Cognition  $F(1,84) = 5.547, p = .021, \eta^2_p = .062$ . However, the unpaid group reported higher occurrences of cognitive failures (113.12, SD = 20.982) versus the paid group (99.34, SD = 18.154),  $F(1,84) = 9.959, p = .002, \eta^2_p = .106$ . This result is consistent with the earlier study in which an unpaid training group reported a greater number of cognitive failures than a paid non-training group (Jaeggi et al., 2014). Participants with higher self-perception of cognitive deficits,

or those who placed greater value on cognitive skills, may have been attracted to the unpaid “brain training” study. This suggests that methods of recruitment -- such as whether participants are compensated or drawn in with the possibility of cognitive improvement -- may be related to certain individual differences across subjects.

Table 4.2. *Baseline trait differences by paid/unpaid condition, using a Bonferroni correction with  $\alpha$  at .00625. Effect size is partial eta-squared.*

	Paid Dual N-back			Unpaid Dual N-back & Knowledge Trainer				
	N	Mean	SD	N	Mean	SD	p	ES
<b>Motivation Factors</b>								
Need for Cognition	36	48.44	35.82	51	67.80	39.00	0.02	0.06
TOCA	36	31.41	7.41	51	33.63	7.66	0.16	0.02
Cognitive Failures	36	99.34	18.15	51	113.12	20.98	0.001*	0.11
<b>Personality Factors</b>								
Extraversion	36	45.00	10.56	51	41.10	12.00	0.11	0.03
Conscientiousness	36	50.28	8.54	51	46.47	10.38	0.07	0.04
Intellectual Openness	36	53.28	7.15	51	53.80	10.55	0.48	0.01
Agreeableness	36	54.22	9.10	51	52.12	8.65	0.31	0.01
Emotional Stability	36	36.25	12.45	51	42.63	10.04	0.01	0.08

**Training Data.** To examine training performance, a slope and intercept were calculated for each participant and the slope was then compared across the three groups. The three training groups (paid dual n-back, unpaid dual n-back, and knowledge trainer) all significantly improved performance on the training task over the four weeks of training ( $ps < .01$ ) (see Figure 4.2 for comparison of the two dual n-back training groups). The paid dual n-back group improved 87%, from an average of 2.32 in the first session to 4.33 in the last session, and the unpaid dual n-back group improved by a similar amount (79%), from an average n-back level of 2.42 in the first session to an n-back level of 4.33 in the last session. The knowledge-trainer group improved by 44% over the course of the training. A comparison of training performance revealed no

significant difference in baseline training performance  $F(1,58) = .273, p = .603, \eta^2_p = .005$ , or the rate of improvement as determined by a linear regression with a logarithmic transformation of the number of session across the paid and unpaid dual n-back training groups  $F(1,58) = .913, p = .343, \eta^2_p = .016$ . A Bayesian follow-up analysis largely supported these results, with  $BF_{10} = 0.299$  for baseline training performance, and  $BF_{10} = 0.396$  for training slope.

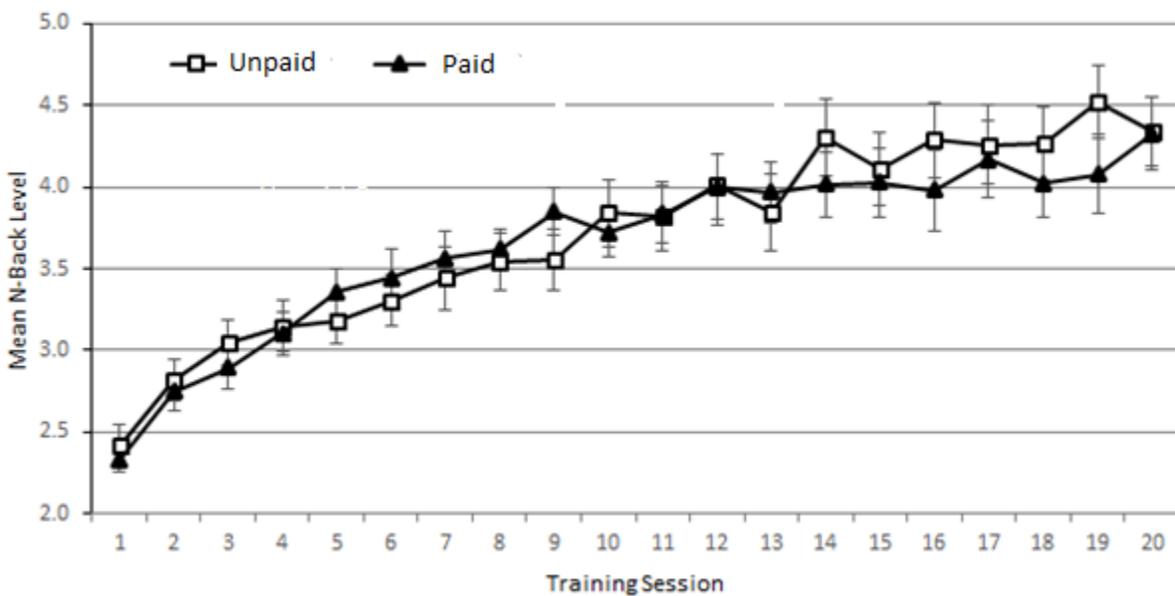


Figure 4.2. Training improvements in the dual n-back training groups.

**Overall Transfer Effects.** Descriptive data and test-retest reliabilities and effect sizes for individual assessments are included in Table 4.3. It should be noted that, for the purpose of calculating the composite scores, z-scores were calculated from the individual pre- and post-test measures (z-scores were not used in Jaeggi et al. (2014)) (see Figure 4.3). There were no significant group differences among pre-test scores of the unpaid dual n-back group, paid dual n-back group, and the knowledge-trainer group for the Visuospatial composite  $F(2,85) = 1.160, p = .318, \eta^2_p = .027$  or the Verbal composite  $F(2,84) = 1.286, p = .282, \eta^2_p = .030$ . Transfer

differences within the present study were assessed with repeated measures ANCOVAs for the pre- and post-test factor scores with initial test version as a covariate (Figure 4.4). Among the paid dual n-back group compared to the knowledge-training group, there was a significant effect of intervention for the Visuospatial Reasoning composite  $F(1,60)=6.737$ ,  $p = .012$ ,  $\eta^2_p = .101$  but not the Verbal Reasoning composite  $F(1,60) = .409$ ,  $p = .525$ ,  $\eta^2_p = .007$ . There were no significant differences in transfer between the paid and the unpaid dual n-back groups, for both the Visuospatial composite  $F(1,58) = .233$ ,  $p = .631$ ,  $\eta^2_p = .004$  and the Verbal composite  $F(1,57) = .274$ ,  $p = .603$ ,  $\eta^2_p = .005$ . Follow-up Bayesian analyses again largely supported both the baseline and transfer results, with  $BF_{10} = .292$  for the Visuospatial composite and  $BF_{10} = .200$  for the Verbal composite at baseline, and with  $BF_{10} = .443$  for the Visuospatial composite and  $BF_{10} = .472$  for the Verbal composite transfer.

Table 4.3. *Individual transfer measures by group*. R = test–retest reliability (partial correlations accounting for test version); ES = effect size that accounts for the correlation between the pre- and post-test measures,  $\frac{\mu_2 - \mu_1}{\sqrt{\sigma_1^2 + \sigma_2^2 - 2r_{12}\sigma_1\sigma_2}}$

	Pretest			Posttest			Pre vs. Post						
	N	Mean	SD	Min	Max	N	Mean	SD	Min	Max	P	R	ES
<b>Dual N-back Paid</b>													
APM	36	13.39	2.38	8	18	36	13.94	2.70	7	18		0.49	0.21
BOMAT	36	16.39	3.86	7	23	36	16.58	3.52	10	23		0.68	0.07
Surface Dev	36	20.00	7.45	7	30	36	22.78	6.22	7	30	**	0.71	0.52
Space Relations	36	12.89	3.43	3	17	36	13.83	2.96	7	17	*	0.66	0.35
Form Board	36	65.00	22.28	24	109	36	78.86	26.44	19	117	***	0.76	0.68
CFT	36	19.47	2.93	13	24	36	19.31	2.96	12	26		0.20	-0.04
Inferences	36	7.72	1.57	2	10	36	7.97	1.63	2	10		0.45	0.14
Reading Com	36	6.06	2.18	1	10	36	6.47	2.26	2	10		0.69	0.24
Verbal Analogies	36	0.78	0.10	0.56	0.91	36	0.80	0.07	0.65	0.93		0.51	0.31
DST	36	72.86	14.20	46	113	36	76.17	15.20	47	113	*	0.79	0.35
<b>Dual N-back Unpaid</b>													
APM	25	15.32	2.63	8	18	25	14.96	2.70	8	18		0.61	-0.15
BOMAT	25	18.24	3.26	12	24	25	18.48	3.90	11	25		0.68	0.08
Surface Dev	25	19.96	8.50	5	30	25	22.88	8.10	4	30	**	0.80	0.55
Space Relations	25	12.40	4.17	4	17	25	13.36	4.08	4	17		0.73	0.32
Form Board	25	61.92	28.20	0	107	25	75.68	25.20	25	123	***	0.77	0.79
CFT	25	19.84	3.95	11	27	25	20.56	3.25	14	26		0.48	0.19
Inferences	25	7.16	1.91	2	10	25	7.68	1.68	3	10		0.31	0.25
Reading Comp	25	7.52	2.02	4	10	25	7.60	2.57	1	10		0.69	0.04
Verbal Analogies	24	0.74	0.10	0.55	0.92	25	0.75	0.08	0.56	0.9		0.40	0.10
DST	24	67.88	13.26	40	88	25	70.36	12.77	42	91		0.86	0.36
<b>Knowledge Training</b>													
APM	27	14.81	2.79	7	18	27	14.74	2.80	8	18		0.60	-0.03
BOMAT	27	18.00	4.10	7	23	27	17.63	3.78	11	26		0.72	-0.12
Surface Dev	27	21.41	8.23	1	30	27	22.56	7.65	0	30		0.84	0.25
Space Relations	27	13.48	3.12	7	17	27	13.81	3.11	4	17		0.43	0.10
Form Board	27	67.41	27.26	5	115	27	66.63	24.49	22	117		0.73	-0.04
CFT	27	19.63	3.12	14	26	27	19.63	3.12	12	24		0.44	0.00
Inferences	27	7.22	1.91	2	10	27	7.96	1.89	2	10	*	0.66	0.47
Reading Comp	27	6.04	2.77	0	10	27	5.85	2.78	1	10		0.74	-0.09
Verbal Analogies	27	0.72	0.10	0.54	0.87	27	0.75	0.09	0.56	0.90		0.29	0.26
DST	27	68.67	11.26	37	91	27	71.85	11.41	57	93	*	0.77	0.41

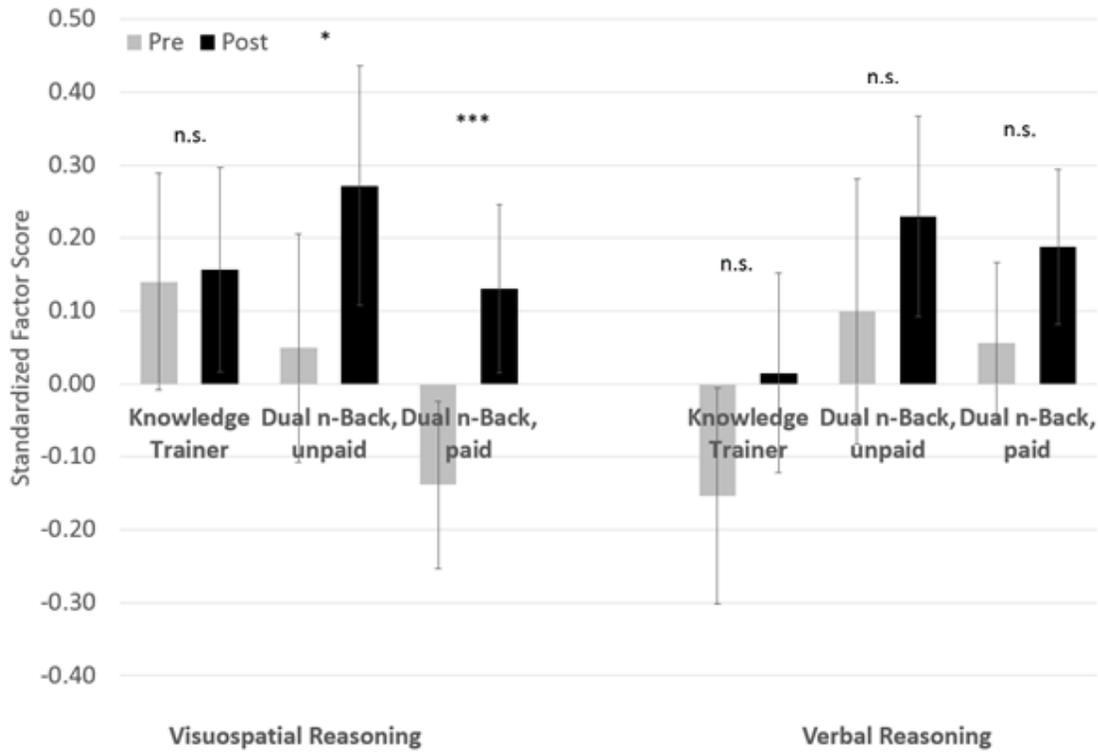


Figure 4.3. Repeated Measures by condition using standardized composite score. Error bars represent the Standard Error of the Mean.

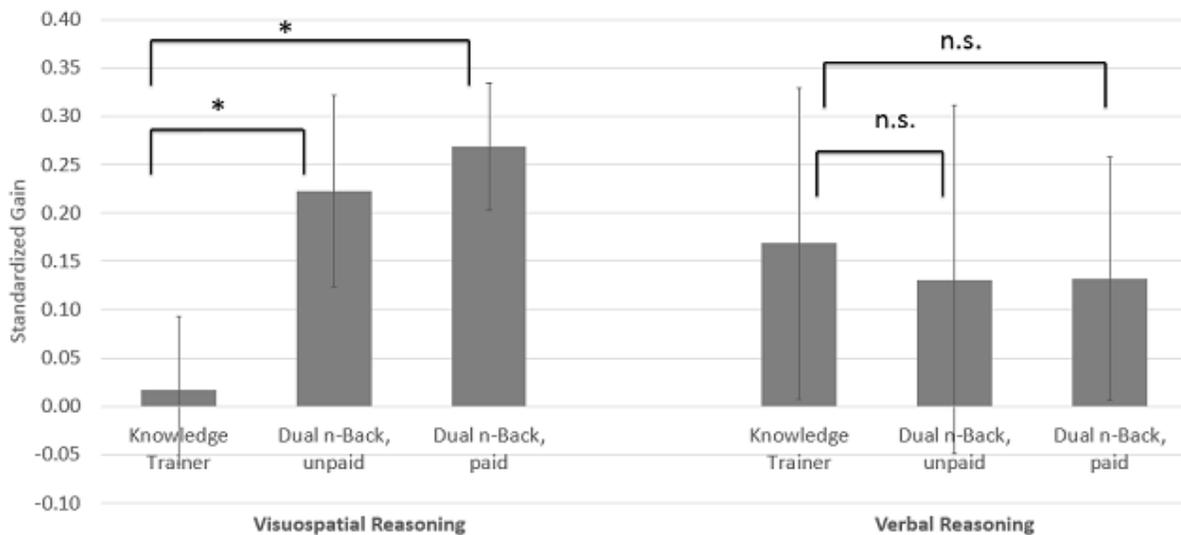


Figure 4.4. Standardized composite gain by dual n-back training group; two-tailed ANCOVA. Error bars represent the Standard Error of the Mean, \*p < .05

The primary purpose of the present research, as a follow-up to a study by Jaeggi et al. (2014), was to determine whether or not motivational factors, specifically the difference between extrinsic motivation related to payment versus intrinsic motivation related to participant-driven desire to improve cognition, might impact transfer gains following cognitive training. As discussed, we identified one common thread across cognitive training studies thus far. That is the possible link between compensation and transfer: studies such as Redick et al. (2013) and Thompson et al. (2013) that provided substantial monetary compensation linked to completion (but not performance) in WM training were generally less successful in demonstrating transfer to visuospatial measures compared those that have not provided significant compensation (Jaeggi et al., 2008; Jaeggi et al., 2010; Stephenson & Halpern, 2013). We hypothesized that compensation might be one determinant of transfer following cognitive training, as previous motivational research has found that monetary compensation may impact performance in linked activities.

The present study found no significant differences in transfer between the compensated and unpaid groups, however. Both improved significantly and by similar amounts on the visuospatial composite, and both experienced no transfer gains on the verbal factor, despite the offer of significant compensation identical to that provided by Redick et al. (2013). While there are many questions regarding the underlying mechanisms of these improvements – for example, it is possible that some other ability besides WM, such as interference resolution, may be improved through the training, and that those improvements may lead to transfer effects – the actual pattern of performance on the outcome measures is quite similar across the paid and unpaid group. As in the experiment by Redick et al., participants in the unpaid group were not recruited using self-improvement language; however, the rest of the protocol mirrored that used in Jaeggi et al. (2014). Several possibilities exist for why the present group experienced transfer

despite significant compensation. One cause may be the inclusion in the consent form of a brief mention that the task may be able to improve intelligence. Given that fifteen individuals later identified this as a motivation for joining the study (despite no improvement language being included in any recruitment materials) it is possible that some participants possessed multiple motivations for completing the research, or that their motivations changed during the month-long training period. This is consistent with motivation research that has found that extrinsic reward, if coupled with self-motivation to complete a task, may not necessarily have a deleterious effect on a targeted activity (Robinson et al., 2012). Additionally, despite not being told that they were part of a “brain-training” study, many participants were probably able to correctly infer that the purpose of the study was to improve cognition. Whether or not participants realize experimenters are attempting to improve fluid intelligence per se, the combination of feedback via adaptivity and the display of training curves likely leads many to conclude that they are being tested on whether they improve on the pre and post measures. Thus, despite our best efforts to hide this purpose from participants in the extrinsic reward group, some intrinsic motivation based on improving cognition may develop across the 20 training sessions.

We divided participants in the paid group into intrinsically motivated ( $N = 15$ ) and extrinsically motivated ( $N = 21$ ) groups depending on their answer to the final motivation check question as given at the post-test. A comparison of these participants at pre-test revealed significant differences in the baseline Visuospatial composite, with higher values for the intrinsically motivated participants ( $3.869$ ,  $SD = .504$ ) versus the extrinsic group ( $3.403$ ,  $SD = .746$ ),  $F(1,34) = 4.204$ ,  $p = .048$ ,  $\eta^2_p = .113$ , but not for the Verbal factor (extrinsic =  $5.11$ ,  $SD = .731$ ; intrinsic =  $5.15$ ,  $SD = .585$ ;  $F(1,34) = .220$ ,  $p = .642$ ,  $\eta^2_p = .007$ ); this result supports the idea that there is a great deal of variability in intrinsic motivation of participants in training

studies regardless of recruitment method. However, there were no significant differences across these groups in training performance  $F(1,34) = .514, p = .478, \eta^2_p = .015$ , or transfer to Visuospatial  $F(1,34) = .791, p = .380, \eta^2_p = .023$  or Verbal composites  $F(1,34) = 1.342, p = .255, \eta^2_p = .039$ . Training curves for the two subsets of the paid dual n-back group are shown in Figure 4.5.

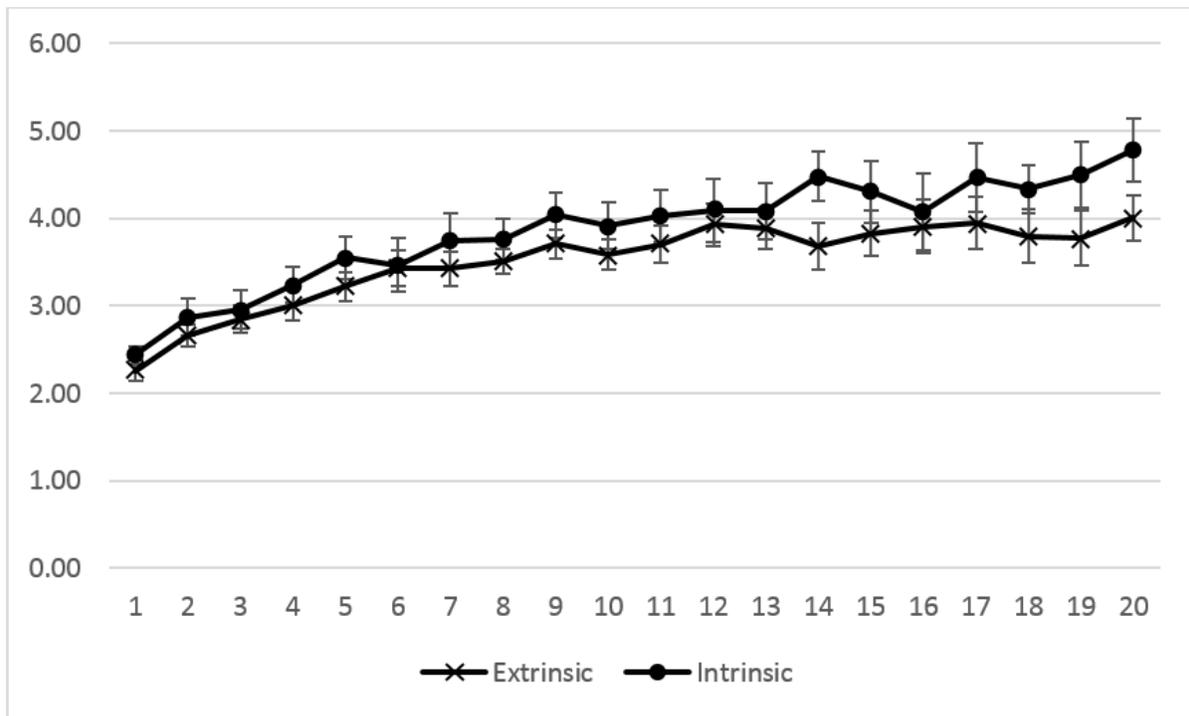


Figure 4.5. Paid group dual n-back training curves as a function of the response to the motivation check question.

Although the small group sizes must temper any conclusions drawn from these data, the varied responses to the final motivation question within the paid dual n-back training group suggests that individuals who sign up for training studies may be differentially motivated to

participate in this research. Despite these individual differences, however, there were again no significant differences in transfer or training gains between these groups.

It is possible that there are a variety of other group differences, such as education level, age, or socioeconomic status that are more closely linked to transfer than compensation, and that may account for the improvements in untrained measures found here as opposed to Redick et al. (2013) and Thompson et al. (2013). In this study a few significant group differences were identified between the paid and unpaid cohort: the compensated dual n-back cohort was younger and reported fewer cognitive failures relative to the unpaid group. Additionally, the attrition rate between the paid and unpaid groups was quite different. In this study the significant compensation may have reduced overall dropout rates from the study, although it is worth noting that at least one previous study with an identical compensation level had a much higher drop-out rate, possibly as a result of other group differences, such as those mentioned above (Redick et al., 2013).

In the present study cohort differences did not impact training improvements or transfer, however. This suggests that, although the addition of extrinsic reward during recruitment may attract different individuals to join a study initially, the nature of these differences and how they contribute to training performance and transfer may be somewhat complex, and, furthermore, may even change throughout the course of a month-long intervention. Within the paid group, the participants who reported only being motivated by money scored lower on the pre-test visuospatial composite than those with an additional intrinsic motivation, suggesting that there may be baseline differences across differently motivated individuals. Again, however, these differences did not significantly impact the outcome of the training or transfer. It is worth considering these differences in light of other studies in which participants were not informed of

potential cognitive improvements from training (Redick et al., 2013), as certain individuals may be more likely to drop out of the study if they would otherwise be self-motivated to continue. Dropout rates for n-back training studies have been so significant across both compensated and uncompensated groups (Jaeggi et al., 2014; Redick et al., 2013) that these considerations must be taken into account in future studies. Compensation, the promise of cognitive improvement in advertising the study, as well as the overall style and tone of the protocol should be more carefully considered when conducting cognitive training research. Nonetheless, the present work demonstrates that compensation need not necessarily negatively impact training improvements and transfer gains.

The similarity between the paid and unpaid group in training and transfer facilitated additional analyses that drew on this increased power and we further examined the effect of personality factors and beliefs on transfer improvement. As discussed earlier, one potential concern is that participants' personal beliefs about the malleability of cognitive abilities may influence training and transfer gains (Jaeggi et al., 2014). If there are differences in beliefs about malleability across different training groups, it is possible that some transfer effects are actually placebo effects. To ensure that the effect of the intervention was significant *even when controlling for participant beliefs about the malleability of cognition*, we computed an ANCOVA comparing the combined dual n-back groups to the knowledge training control group, with gain on the visuospatial composite as the dependent variable and participant score on the Theories of Cognitive Abilities Scale (TOCA) as a covariate. This analysis found that TOCA accounted for a marginal portion of the variance  $F(1,83) = 3.260, p = .075, \eta p^2 = .038$ . Even when controlling for TOCA, the comparison of the combined dual n-back groups versus the knowledge trainer group revealed that the effect of the intervention remained significant,

suggesting that there was little to no placebo effect  $F(1,83)=6.003$ ,  $p = .016$ ,  $\eta p^2 = .067$ .

Nevertheless, it remains possible that participants' personal beliefs impact the outcome of the training in meaningful ways that we have not investigated here; for example, it is possible that individuals may have different expectations about the effectiveness of different types of training. That is, participants may believe that the visuospatial aspects of the dual n-back task are more likely to improve visuospatial abilities compared to the knowledge trainer, which is entirely verbal. To conclusively test for different possible placebo effects, larger studies with multiple control groups would be required; the goal of this study, however, is not to resolve questions about the effectiveness of cognitive training but to examine the impact of monetary compensation on these interventions.

**Conclusion.** The current study finds that even a relatively substantial amount of monetary compensation for participation does not seem to impact the outcome of WM training. This finding was unexpected considering the extant literature and the recent meta-analysis that investigated compensation and cognitive training (Au et al., 2014). However, even the Au et al. meta-analysis did not find a significant effect of remuneration on the outcome of training once an outlier study was removed from the analysis. One potential reason why, regardless of compensation, that we found transfer in this study is that our protocol may have ensured that our participants maintained strong intrinsic motivation over the course of the intervention. A classic study of motivation finds that the actual language and style of the protocol may matter a great deal in alleviating any undermining effect payment may have—especially if that language encourages participants to perform well (Ryan, Mims, & Koestner, 1983).

These results suggest that remuneration may be not be harmful in cognitive training studies.

Compensation may even aid in compliance with the training, as the attrition rate from the paid

dual n-back group was significantly lower than that in the unpaid group. Even participants who did not state an intrinsic motivation for signing up for the study may nonetheless have engaged more during the training program due to a sense of personal improvement during the course of training. Therefore we suggest that monetary compensation, if offered in a careful fashion that does not interfere with participants' intrinsic motivation to improve cognition and apply themselves to the training paradigm, may help future researchers in motivating participants to complete these lengthy and time-intensive training regimens. Future studies should help to establish which individual differences in motivation may ultimately influence transfer improvements following cognitive training, as well as whether specific protocol and methodological features may help or hinder these outcomes.

It is also interesting to note that, unlike Foroughi et al (2016), we did not find a significant difference between the control and experimental group based on the Theories of Intelligence Scale. Furthermore, there was no difference in performance on the outcome of training between the two groups, even though our recruitment flyers differed in a similar manner to those used in the Foroughi study. It may be that while placebo effects may be present throughout training to some degree, they only have a meaningful impact on the outcome of training in a very short intervention. Because most cognitive training studies include many sessions of training, placebo effects may only rarely play a role; however, like Foroughi and colleagues, we recommend that future studies examine placebo effects to learn whether this is indeed the case.

These results have important implications for future research in WM training, but several issues limit the conclusions that may be drawn from the present study. The paid and unpaid cohorts were collected in different semesters so that no participant would respond to both

posters; this may be a factor contributing to some demographic differences between these groups. Additionally, the knowledge-training comparison group was collected only as part of the unpaid cohort; a matched-paid group was not examined as part of the study. Despite these differences, the pattern of transfer across the two dual n-back groups was remarkably similar. This result suggests that the effects of dual n-back training may be more resilient to certain moderating factors than previously expected -- at least under conditions in which the experimental protocol is carefully implemented.

### References

- Amabile, T. M., Hill, K. G., Hennessey, B. A., & Tighe, E. M. (1994). The Work Preference Inventory: assessing intrinsic and extrinsic motivational orientations. *Journal of Personality and Social Psychology*, 66(5), 950.
- Anguera, J. A., Bernard, J. A., Jaeggi, S. M., Buschkuhl, M., Benson, B. L., Jennett, S., Humfleet, J., Reuter-Lorenze, P. A., Jonides, J., & Seidler, R. D. (2012). The effects of working memory resource depletion and training on sensorimotor adaptation. *Behavioural Brain Research*, 228 (1), 107-115.
- Au, J., Sheehan, E., Tsai, N., Duncan, G. J., Buschkuhl, M., & Jaeggi, S. M. (2014). Improving fluid intelligence with training on working memory: a meta-analysis. *Psychonomic Bulletin & Review*, 1-12.
- Bennett, G. K., Seashore, H. G., & Wesman, A. G. (1972). *Differential Aptitude Test: Space Relations*. New York: Psychological Corporation.
- Benware, C. A., & Deci, E. L. (1984). Quality of learning with an active versus passive motivational set. *American Educational Research Journal*, 21(4), 755-765.
- Berger, F. R., Gupta, W. B., & Berger, R. M. (1990). *Air Force officer qualifying test (AFOQT) Form P: Test manual*. Psychometrics Inc, Sherman Oaks CA.
- Blumenfeld, P. C., Kempler, T. M., & Krajcik, J. S. (2006). *Motivation and Cognitive Engagement in Learning Environments*. New York: Cambridge University Press.
- Cacioppo, J. T., & Petty, R. E. (1982). The Need for Cognition. *Journal of Personality and Social Psychology*, 42(1), 116-131.
- Cameron, J., & Pierce, W. D. (1994). Reinforcement, reward, and intrinsic motivation: A meta-analysis. *Review of Educational Research*, 64(3), 363-423.

- Cattell, R. B., & Cattell, A. K. S. (1963). *Test of "g": Culture Fair Scale 3*. Champaign, IL: Institute for Personality and Ability Testing.
- Chooi, W. T., & Thompson, L. A. (2012). Working memory training does not improve intelligence in healthy young adults. *Intelligence, 40*(6), 531-542.
- Colom, R., Román, F. J., Abad, F. J., Shih, P. C., Privado, J., Froufe, M., Escorial, S., Martinezm K., Burgeleta, M., Quiroga, M. A., Karama, S., Haier, R. J., Thompson, P. M., & Jaeggi, S. M. (2013). Adaptive n-back training does not improve fluid intelligence at the construct level: Gains on individual tests suggest that training may enhance visuospatial processing. *Intelligence, 41*(5), 712-727.
- Condry, J., & Chambers, J. (1978). Intrinsic motivation and the process of learning. In M. R. Lepper & D. Greene (Eds.), *The hidden costs of reward: New perspectives on the psychology of human motivation*, 61-84. Hillsdale, N.J: Erlbaum.
- Deci, E. L. (1971). Effects of externally mediated rewards on intrinsic motivation. *Journal of personality and Social Psychology, 18*(1), 105.
- Deci, E. L. (1972). The effects of contingent and noncontingent rewards and controls on intrinsic motivation. *Organizational Behavior and Human Performance, 8*(2), 217-229.
- Deci, E. L., Koestner, R., & Ryan, R. M. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological bulletin, 125*(6), 627.
- Dev, P. C. (1997). Intrinsic Motivation and Academic Achievement What Does Their Relationship Imply for the Classroom Teacher?. *Remedial and Special Education, 18*(1), 12-19.
- Durik, A. M., & Harackiewicz, J. M. (2007). Different strokes for different folks: How individual interest moderates the effects of situational factors on task interest. *Journal of Educational Psychology, 99*(3), 597.
- Dweck, C. S. (1999). *Self-Theories: Their role in motivation, personality, and development*. Philadelphia, PA: Psychology Press.
- Dweck, C. S., & Master, A. (2008). Self-theories motivate self-regulated learning. In D. H. Schunk & B. J. Zimmerman (Eds.), *Motivation and self-regulated learning: Theory, research, and applications*, 31-51. New York, NY: Erlbaum.
- Ekstrom, R. B., French, J. W., Harmon, H. H., & Derman, D. (1976). *ETS Kit of Factor-Referenced Cognitive Tests*. Princeton, NJ: Educational Testing Service.
- Henderlong, J., & Lepper, M. R. (2002). The effects of praise on children's intrinsic motivation: a review and synthesis. *Psychological Bulletin, 128*(5), 774.

- Hossiep, R., Turck, D., & Hasella, M. (1999). *Bochumer Matrizentest. BOMAT advanced*. Göttingen: Hogrefe.
- Foroughi, C. K., Monfort, S. S., Paczynski, M., McKnight, P. E., & Greenwood, P. M. (2016). Placebo effects in cognitive training. *Proceedings of the National Academy of Sciences of the USA*, 201601243.
- Fransson, A. (1977). On qualitative differences in learning: IV—Effects of intrinsic motivation and extrinsic test anxiety on process and outcome. *British Journal of Educational Psychology*, 47(3), 244-257.
- Jaeggi, S. M., Buschkuhl, M., Jonides, J., & Perrig, W. J. (2008). Improving fluid intelligence with training on working memory. *Proceedings of the National Academy of Sciences of the USA*, 105(19), 6829-6833.
- Jaeggi, S. M., Buschkuhl, M., Jonides, J., & Shah, P. (2011). Short-and long-term benefits of cognitive training. *Proceedings of the National Academy of Sciences of the USA*, 108(25), 10081-10086.
- Jaeggi, S. M., Buschkuhl, M., Shah, P., and Jonides, J. (2014). The role of individual differences in cognitive training and transfer. *Memory & Cognition*. 42(3), 464–480.
- Jaušovec, N., & Jaušovec, K. (2012). Working memory training: improving intelligence—changing brain activity. *Brain and Cognition*, 79(2), 96-106.
- Johnson, W., te Nijenhuis, J., & Bouchard, T. J. (2008). Still just 1 g: Consistent results from five test batteries. *Intelligence*, 36(1), 81-95.
- Judge, T. A., & Ilies, R. (2002). Relationship of personality to performance motivation: a meta-analytic review. *Journal of Applied Psychology*, 87(4), 797.
- Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working memory capacity: a latent-variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology: General*, 133(2), 189-217.
- Kane, M. J., & Miyake, T. M. (2007). The validity of "conceptual span" as a measure of working memory capacity. *Memory & Cognition*, 35(5), 1136-1150.
- Kundu, B., Sutterer, D. W., Emrich, S. M., & Postle, B. R. (2013). Strengthened effective connectivity underlies transfer of working memory training to tests of short-term memory and attention. *The Journal of Neuroscience*, 33(20), 8705-8715.
- Lin, Y. G., McKeachie, W. J., & Kim, Y. C. (2001). College student intrinsic and/or extrinsic motivation and learning. *Learning and Individual Differences*, 13(3), 251-258.

- McVay, J. C., & Kane, M. J. (2009). Conducting the train of thought: Working memory capacity, goal neglect, and mind wandering in an executive-control task. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 35(1), 196-204.
- Morrison, A. B., & Chein, J. M. (2011). Does working memory training work? The promise and challenges of enhancing cognition by training working memory. *Psychonomic Bulletin & Review*, 18(1), 46-60.
- Murayama, K., Matsumoto, M., Izuma, K., & Matsumoto, K. (2010). Neural basis of the undermining effect of monetary reward on intrinsic motivation. *Proceedings of the National Academy of Sciences of the USA*, 107(49), 20911-20916.
- Pintrich, P. R. (1999). The role of motivation in promoting and sustaining self-regulated learning. *International Journal of Educational Research*, 31(6), 459-470.
- Pittman, T. S., & Heller, J. F. (1987). Social motivation. *Annual Review of Psychology*, 38(1), 461-490.
- Prins, P. J., Doyis, S., Ponsioen, A., Ten Brink, E., & Van der Oord, S. (2011). Does computerized working memory training with game elements enhance motivation and training efficacy in children with ADHD? *Cyberpsychology, Behavior, and Social Networking*, 14(3), 115-122.
- Raven, J. C. (1990). *Advanced Progressive Matrices. Sets I, II*. Oxford: Oxford University Press.
- Redick, T. S., Shipstead, Z., Harrison, T. L., Hicks, K. L., Fried, D. E., Hambrick, D. Z., & Engle, R. W. (2013). No evidence of intelligence improvement after working memory training: A randomized, placebo-controlled study. *Journal of Experimental Psychology: General*, 142(2), 359.
- Reeve, J., & Deci, E. L. (1996). Elements of the competitive situation that affect intrinsic motivation. *Personality and Social Psychology Bulletin*, 22, 24-33.
- Robinson, L. J., Stevens, L. H., Threapleton, C. J., Vainiute, J., McAllister-Williams, R. H., & Gallagher, P. (2012). Effects of intrinsic and extrinsic motivation on attention and memory. *Acta psychologica*, 141(2), 243-249.
- Rudebeck, S. R., Bor, D., Ormond, A., O'Reilly, J. X., & Lee, A. C. (2012). A potential spatial working memory training task to improve both episodic memory and fluid intelligence. *PLoS One*, 7(11), e50431.
- Ryan, R. M., Mims, V., & Koestner, R. (1983). Relation of reward contingency and interpersonal context to intrinsic motivation: A review and test using cognitive evaluation theory. *Journal of Personality and Social Psychology*, 45(4), 736.

- Saucier, G. (1994). Mini-markers: A brief version of Goldberg's unipolar Big-Five markers. *Journal of Personality Assessment*, 63(3), 506-516.
- Schweizer, S., Hampshire, A., & Dalgleish, T. (2011). Extending brain-training to the affective domain: increasing cognitive and affective executive control through emotional working memory training. *PloS One*, 6(9), e24372.
- Spence, J. T., & Helmreich, R. L. (1983). Achievement-related motives and behaviors. In J. T. Spence (Ed.). *Achievement and achievement motives: Psychological and sociological approaches*, 7-74. San Francisco, CA: W.H. Freeman and Company.
- Stephenson, C. L., & Halpern, D. F. (2013). Improved matrix reasoning is limited to training on tasks with a visuospatial component. *Intelligence*, 41(5), 341-357.
- Studer-Luethi, B., Jaeggi, S. M., Buschkuhl, M., & Perrig, W. J. (2012). Influence of neuroticism and conscientiousness on working memory training outcome. *Personality and Individual Differences*, 53(1), 44-49.
- Swann Jr, W. B., & Pittman, T. S. (1977). Initiating play activity of children: The moderating influence of verbal cues on intrinsic motivation. *Child Development*, 1128-1132.
- Tang, S. H., & Hall, V. C. (1995). The overjustification effect: A meta-analysis. *Applied Cognitive Psychology*, 9(5), 365-404.
- Thompson, T. W., Waskom, M. L., Garel, K. L. A., Cardenas-Iniguez, C., Reynolds, G. O., Winter, R., & Gabrieli, J. D. (2013). Failure of working memory training to enhance cognition or intelligence. *PloS One*, 8(5), e63614.
- Vollmeyer, R., & Rheinberg, F. (2000). Does motivation affect performance via persistence? *Learning and Instruction*, 10(4), 293-309.
- Wechsler, D. (1997). *Wechsler Adult Intelligence Scale* (3rd ed.). San Antonio, TX: Psychological Corporation.
- Wiersma, U. J. (1992). The effects of extrinsic rewards in intrinsic motivation: A meta-analysis. *Journal of Occupational and Organizational Psychology*, 65(2), 101-114.
- Wright, R., Thompson, W. L., Ganis, G., Newcombe, N. S., & Kosslyn, S. M. (2008). Training generalized spatial skills. *Psychonomics Bulletin & Review*, 15(4), 763-771.
- Zinke, K., Zeintl, M., Rose, N. S., Putzmann, J., Pydde, A., & Kliegel, M. (2014). Working memory training and transfer in older adults: Effects of age, baseline performance, and training gains. *Developmental Psychology*, 50(1), 304.

## CHAPTER 5

### Individual Differences and tDCS-Augmented Cognitive Training

#### Introduction

Considering the importance of working memory (WM) for success in a wide variety of real-life contexts, including school (Alloway & Alloway, 2010) and work (Higgins, Peterson, Pihl, & M, 2007), it is unsurprising that a variety of WM interventions have been proposed in recent years. Transcranial direct current stimulation (tDCS) and cognitive training are two cognitive enhancement techniques that have recently been used together to improve WM, with promising, but by no means conclusive, results. A recent meta-analysis from Mancuso et al. (2016) suggests that dorsolateral prefrontal cortex (DLPFC) stimulation during training results in a small but significant enhancement effect, which survives corrections for publication bias. Recent research (Au et al., 2016) provides further evidence that DLPFC stimulation (both right and left) enhances performance on a widely used *n*-back training task over the course of seven sessions, relative to a Sham stimulation condition. While these initial findings do provide some preliminary support for the use of tDCS to enhance learning of WM-intensive tasks, we note considerable heterogeneity in the literature. For example, a similarly designed *n*-back/tDCS training study failed to find an effect of tDCS after correcting for baseline differences (Martin et al., 2013), and the ten tDCS/WM training studies covered in the Mancuso et al. (2016) meta-analysis differ substantially in the magnitude of their effects, with Hedges' *g* values ranging from 0.074 to 0.565. A variety of factors, including differences in stimulation intensity, density,

location, and other parameters, as well as the design and implementation of the cognitive training paradigm, may explain the disparities in the strength of these effects (see Au et al., 2016 for a brief discussion). However, one additional possibility is that individual differences among participants – including motivation, gender, and baseline ability, among many factors – may play important roles. These factors may influence the outcome of the combined intervention in their own right, but they may also be associated with other individual-difference characteristics that influence performance (for example, different geographic training locations may be confounded with educational background). While extant research does suggest that individual differences play a significant role in both tDCS interventions (Krause & Cohen Kadosh, 2014) and cognitive training interventions (Jaeggi, Buschkuhl, Shah, & Jonides, 2014; Katz, Jones, Shah, Buschkuhl, & Jaeggi, in press) by themselves, these factors have rarely been investigated directly in studies that combine both interventions.

**Baseline performance and other individual-difference factors in tDCS.** Studies by Wiethoff, Hamada, and Rothwell (2014) and Lopez-Alonso et al. (2014) have found that even in tDCS experiments that successfully demonstrate an effect on cognition overall, less than half of the participants demonstrate improved performance. This suggests that a considerable proportion of participants in each study may not be responding to the treatment. Additionally, recent work has raised controversy about the previously dominant neural explanation for tDCS-related cognitive enhancement (Buzsáki, 2016). While the consensus thus far has been that anodal stimulation causes depolarization of the resting membrane potential, facilitating the production of action potentials, Buzsáki's work with cadavers questions the amount of current that actually reaches the cortex. Thus it is possible that certain individual physical characteristics could have a larger effect than expected previously. For example, even something as seemingly minor as hair

thickness may impact electrode contact and further reduce the amount of current passing through the scalp and skull. However, several individual-difference factors have been studied in conjunction with tDCS prior to Buzsáki's provocative findings. Kraus and Cohen Kadosh (2014) suggested that age, gender, and neuronal factors, namely regional cortical excitability, may influence the effectiveness of transcranial electrical stimulation. For example, it has been proposed that an optimal balance of excitation/inhibition in different cortical regions promotes optimal cognitive functioning. Therefore, tDCS may exert different and sometimes contradictory effects in populations that vary with respect to this balance, such as those with ADHD or depression (Krause, Marquez-Ruiz, & Cohen Kadosh, 2013). Furthermore, genetic factors (Brunoni et al., 2013; Plewnia et al., 2013) and anatomical differences that impact the electric field generated by tDCS (Kim et al., 2014) may also influence the response to stimulation. In addition to these physiological characteristics, it is also possible that psychological characteristics, such as baseline cognitive ability, may influence the outcome of stimulation. Several studies have demonstrated selective tDCS benefits among individuals with low, but not high, baseline WM abilities (Gozenman & Berryhill, 2016; Heinen et al., 2016; Tseng et al., 2012), and meta-analyses tend to report stronger effect sizes in clinical or older adult populations compared to the higher-performing young adult population (Dedoncker, Brunoni, Baeken, & Vanderhasselt, 2016; Hill, Fitzgerald, & Hoy, 2016; Hsu, Ku, Zanto, & Gazzaley, 2015; Summers, Kang, & Cauraugh, 2015). Moreover, the evidence extends beyond the WM domain. Individuals with poorer motor coordination (McCambridge, Bradnam, Stinear, & Byblow, 2011; Uehara, Coxon, & Byblow, 2015), postural control (Zhou et al., 2015), visual acuity (Reinhart, Xiao, McClenahan, & Woodman, 2016), and attention (London & Slagter, 2015; Sikstrom et al., 2016) all showed improvement in the relevant domains while their higher-performing peers did

not. However, it should be noted that these low-baseline effects are not found universally. One group of researchers has repeatedly found an advantage for high-baseline individuals on WM performance during parietal stimulation (Berryhill & Jones, 2012; Jones & Berryhill, 2012; Jones, Gozenman, & Berryhill, 2015), which has been replicated by others (Learmonth, Thut, Benwell, & Harvey, 2015). Another group examining lateralized attention bias found both high- and low-baseline advantages in two separate experiments, but the direction of this advantage depended critically on stimulation intensity (Benwell, Learmonth, Miniussi, Harvey, & Thut, 2015). Therefore, there is no consensus on the influence of baseline performance at present. Also, there are likely even more nuanced issues to consider, such as the brain region stimulated and task-specific optimum levels of neural activity. Thus, there is considerable value in studying tDCS effects in the context of baseline ability, as well as other individual-difference factors.

**Baseline performance and other individual-difference factors in working memory training.** Some research has also been done to examine the effects of individual-difference factors in the outcome of WM training by itself, unaided by tDCS. For example, baseline performance has also been studied in this context, and much like in the tDCS literature, there is also evidence that baseline WM abilities could impact training performance in two possible directions. Some have suggested that individuals with a lower baseline score should have more room to improve at the trained task during the intervention; for example, Zinke and colleagues have demonstrated this through two studies with older adults (Zinke, Zeintl, Eschen, Herzog, & Kliegel, 2012; Zinke et al., 2014). Others have posited that individuals with higher baseline WM performance are better prepared to take advantage of the intervention and thus improve more throughout the training (Lövdén, Bäckman, Lindenberger, Schaefer, & Schmiedek, 2010). There is not yet consensus regarding the impact of baseline performance for the outcome of cognitive

training; it also remains possible that ceiling effects and differences in the design of the training intervention itself may also influence the relationship between starting WM ability and level of improvement in any individual study.

A variety of other individual-difference factors have also been discussed in the context of cognitive training. For example, motivation to complete a task may influence how receptive one is to a training intervention (Jaeggi, Buschkuhl, Jonides, & Shah, 2011; Jaeggi et al., 2014). Many interventions include game-like elements that may influence a participant's motivation as well as his/her performance on the task (Katz, Jaeggi, Buschkuhl, Stegman, & Shah, 2014; Prins, DAVIS, Ponsioen, ten Brink, & van der Oord, 2010), with the general expectation that more motivated or engaged participants will perform better in the training. Additionally, many training studies provide considerable monetary remuneration for participation, and it is possible that this payment may undermine motivation and thus impact overall performance (Au et al., 2015). As mentioned earlier, the study location (e.g., university vs. small college but also different countries; c.f., Au et al., 2015) may influence the outcome of training, although it is difficult to identify which element of geographic location, including cultural factors, actually may play a role in performance. Age has also been studied extensively as a factor that may determine performance on cognitive training tasks. In general, older individuals seem to improve less on untrained tasks administered at pre- and post-test, as well as on the training task itself (Borella et al., 2014; Brehmer, Westerberg, & Bäckman, 2012; Schmiedek, Lovden, & Lindenberger, 2010; Zinke et al., 2014). Although one meta-analysis found no differences in transfer improvements as a function of age (Karbach & Verhaeghen, 2014), another meta-analysis with a larger range of ages found that younger adults improved more on untrained tasks than older adults (Wass, Scerif, & Johnson, 2012). These age-related disparities make some sense given well-established

differences in age-related WM performance (D. C. Park et al., 2002) and theoretical perspective on cognitive plasticity and aging (Lövdén et al., 2010). However, it remains unknown whether age-related differences in cognitive training performance are due to differences in baseline performance or other factors related to aging. Finally, traits such as conscientiousness and neuroticism (Barbara Studer-Luethi, Bauer, & Perrig, 2015; B. Studer-Luethi, Jaeggi, Buschkuhl, & Perrig, 2012) may also impact the outcome of training, while other factors, such as gender, have been found to influence the outcome of training in some studies (Söderqvist, Bergman Nutley, Ottersen, Grill, & Klingberg, 2012) but not others (Klingberg et al., 2005). It remains possible that a number of other factors that have been largely unexplored (e.g., socioeconomic status, although see Segretin et al., 2014) may play a role, at least in some interventions.

Given the relevance of individual-difference factors to the outcome of cognitive training and tDCS independently, a salient question is how these individual-difference factors influence combined interventions featuring both tDCS and WM training together. It is possible, and perhaps even likely, that there are interactions between these two interventions such that some individual-difference factors matter more than others, particularly in the outcome of a combined intervention. For example, in light of the evidence that baseline cognitive ability impacts both the amount one is able to improve during a training intervention and the participant's response to tDCS, it is possible that it will play a much larger role in a combined intervention. The relative paucity of tDCS-augmented cognitive training studies means that it is unsurprising that these factors have not yet been explored in combined interventions. However, given the possibility that they may play a substantial role in the outcomes of such interventions, there is considerable impetus for studying them. Thus, the present paper uses a recently published dataset of tDCS and

WM training data to evaluate the influence of individual differences including baseline performance, motivation, gender, and geographic training location on WM training performance.

As illustrated above, individual differences are one topic of relevance in improving our understanding of why stimulation-augmented cognitive training may be effective for any individual participant. Another point of significant practical importance is how durable training improvements may be over the weeks and months following the intervention. It would likely not make sense to utilize tDCS/WM interventions in real-world applications if the improvements generated by the stimulation dissipated shortly after the intervention. While previous research suggests that there is durability even several months following the intervention (Au et al., 2016), little extant tDCS work examines the stability of improvements over time, and results from WM training research suggest that washout may be a common occurrence within a short time following a training intervention (Melby-Lervåg & Hulme, 2013). By contrast, some studies suggest that improvements following tDCS interventions may remain weeks or even months following the stimulation. Jeon et al. (2012), Jones et al. (2015), and Park et al. (2014) all found continued improvements to WM performance from a week to two months following stimulation. Persistent, long-term changes have also been detected as a function of learning or training in other domains as well, such as motor-skill training (Reis et al., 2009), math training (Looi et al., 2016), and episodic-memory retrieval (Manenti, Sandrini, Brambilla, & Cotelli, 2016). However, to our knowledge, no other study of combined tDCS and cognitive training has examined whether these follow-up effects are maintained for time periods in excess of two to three months after the intervention. In the present study we added to the follow-up findings from Au et al. (2016), including new data not previously reported in which participants returned an average of

12 months following the intervention to complete one more session of the WM training (without stimulation).

## **Methods**

**Participants.** Our dataset comprised largely the same participants as that of Au et al. (2016), which recruited healthy, right-handed individuals between the ages of 18 and 35 as part of a collaborative effort from the campuses of the University of California, Irvine (UCI) and the University of Michigan, Ann Arbor (UM). Six additional individuals completed study procedures, one of whom was excluded as an outlier (see Results), for a total sample size of 67 in the current dataset. As before, participants were excluded if they had had any history of psychological or neurological disorders (including seizures or strokes), previous cognitive training or neurostimulation, past or present drug/alcohol abuse, or if they were taking any medications that would affect attention or memory. All research procedures were approved by the Institutional Review Boards at both universities, and each participant provided informed consent.

**General Procedure.** The experiment, an extension of a previous report (Au et al., 2016), consisted of a between-subjects pretest-posttest intervention design in which participants were randomized into one of two groups. Forty received Active tDCS (Active group) over the right or left DLPFC and 27 received Sham stimulation (Sham group) to the same regions in which current was turned off after the first 30 seconds without the participants' knowledge. The previous report analyzed the right and left DLPFC groups separately in the Active condition, but since we found no differences in the training effect, they are collapsed together in the present chapter. Both groups completed seven days of visuospatial n-back training concurrently with either Active or Sham stimulation.

After the initial training period, all participants were invited back for two follow-up sessions to examine the stability of training effects. Forty-one participants returned for the first follow-up (27 Active and 14 Sham), as reported previously (Au et al., 2016), and 26 participants returned for the second follow-up in the present study (18 Active and 8 Sham). Due to the long delay, the follow-up visits were marred by substantial attrition rates, but 25 of the 26 participants in the second follow-up also participated in the first follow-up, thereby allowing us to evaluate the longitudinal trajectory of a stable cohort of individuals. The mean delay after the initial training period was 221 days (range: 97 – 393; SD = 82) for the first follow-up and 355 days (range: 251 - 471; SD=73) for the second follow-up. Maintenance of transfer effects was not evaluated at this second follow-up due to the lack of sustained transfer during the first follow-up.

**Working Memory Training.** The training task was a computerized adaptive visuospatial n-back task in which a series of blue squares was displayed one at a time, each in one of eight possible spatial locations. Participants were asked to indicate whether the current square was in the same position as the square presented  $n$  trials ago by responding with the letter “A” to targets and “L” to non-targets, using a standard computer keyboard. The difficulty of the task adapted continuously based on the trainee’s performance. The average n-back level at which a participant trained was calculated each day, and the primary dependent variable for analysis was the logarithmic slope of the 7-session training curve. Further details regarding the design of the training task can be found in Au et al. (2016).

**Transcranial Direct Current Stimulation.** Stimulation was administered via a Soterix Medical 1x1 Low-Intensity tDCS device (Model 1300A) using 5x7cm sponge electrodes placed horizontally on the head. The anode was placed over either right or left DLPFC (sites F4 and F3 in the international 10-20 EEG system) and the cathode was placed over the contralateral

supraorbital area (sites Fp1 or Fp2). Stimulation lasted 25 minutes, with a current intensity of 2mA, which ramped up and down for the first and last 30 seconds of stimulation. Sham tDCS was set up in exactly the same way, except that the current was shut off between the 30-second ramping periods at the beginning and end of each session.

**Individual-difference Variables.** *Baseline:* A baseline score for each participant was determined by his/her visual n-back score at pre-test, measured as  $P_r$ , the proportion of hits minus the proportion of false alarms (Snodgrass & Corwin, 1988). The visual n-back task, which required participants to identify whether a series of colored balls matched the color of the items presented  $n$  before, is similar but not identical to the trained visuospatial n-back, which involved sequential presentation of a square in different spatial locations. In the absence of a true unstimulated baseline of the actual training task, the visual n-back was chosen as the closest reasonable proxy. Although our pretest battery consisted of four WM tasks—visual n-back, auditory n-back, digit span, and Corsi blocks—the latter two are span tests which correlate only weakly with n-back performance (Redick & Lindsey, 2013), while the former two are structurally similar to the trained visuospatial n-back. Although the previous report (Au et al., 2016) made the a priori decision to combine these two n-back tests into a composite measure to test for group differences in baseline, we ended up finding strong transfer effects only in the visual, but not auditory, n-back test. This suggests a close link between visual n-back and our visuospatial training task (correlation between pre-test visual n-back score and first visuospatial n-back training session:  $r = .65$ ). Therefore, it is chosen as the most appropriate index of baseline WM ability in the current study. The average baseline performance on the visual n-back task in the Active group was .66 ( $SD=.16$ ) and the average score in the Sham group was .62 ( $SD=.19$ ); the difference between groups was not significant ( $p = .36$ ).

*Motivation:* Motivation was assessed before each training session by self-report.

Participants were asked to rate their own motivational state on a scale from 1-10, with 10 being the most highly motivated. An average motivation score over all seven sessions was calculated for each participant and used as the dependent variable in analyses. Average motivation scores were 6.1 (SD=1.24) in the Active group and 6.1 (SD=1.01) in the Sham group. We note that although motivation was briefly considered in the previous report (Au et al., 2016), the analysis focused only on confirming the stability of motivation across groups and time, and we did not previously evaluate motivation as an individual-difference factor to predict training outcome as in the current chapter.

*Gender:* Gender information was collected as part of a standard demographic questionnaire during the consent process. The Active group comprised 60% women, while the Sham group comprised 67% women.

*Training Site:* 50% of Active participants were recruited on each campus (UCI and UM), while 59% of Sham participants were recruited at UM.

**Analytic Approach.** Statistical analyses were conducted using IBM SPSS Statistics version 22 and STATA version 13. To identify the effects of individual-difference variables on training performance, separate regression models were calculated for each variable of interest using parameters of a logarithmic model run on the training data, yielding a 7-session training slope as the outcome variable, with condition, the variable of interest, and their interaction as prediction terms. Note that Au et al. (2016) used a seven-level repeated measures ANOVA to analyze training performance. However, for our current analyses, we required an index of training performance as an outcome variable for the regression models. We opted for individual slopes in order to take into account the entire trajectory of training performance. Individual-

difference variables included gender, school site, motivation, and baseline *n*-back performance. All continuous variables were standardized, and thus also mean-centered, while categorical variables remain unstandardized to preserve the inherent structure of the dummy coding and to maintain interpretability. To identify the effects of the long-term follow-up, a similar method was used as in Au et al. (2016) in which gain on the training task was calculated by subtracting performance in the follow-up session from that of the initial training session. This gain was then used as the dependent variable in an ANCOVA with condition as a between-subjects factor. Due to the post-hoc nature of this follow-up, the time lag between the final session of the initial intervention and the follow-up varied between participants and thus was included as a covariate.

## Results

**Outlier Analysis.** Outliers in the data were evaluated by examining the average training performance across all 7 sessions for each participant, as done previously (Au et al., 2016). Outliers were only examined in the Sham group because no new Active participants were enrolled since the previous report. Using a criterion of 2 SD, we identified one high-performing outlier who trained at an average *n*-back level of 7.9, almost twice the group average of the remaining Sham participants (mean: 4.19, SD=1.27). However, we also note that the primary findings presented below are not impacted by the presence or absence of this outlier.

**Training Performance by Condition (Active vs. Sham).** Because five participants were added to the sample beyond the participants included in Au et al. (2016), and because here we use the parameters of a logarithmic model (slope of training curve) as a measure of training progress (instead of the mixed ANOVAs with training performance for each session as used before), an initial model was calculated to re-establish the difference between the Sham and

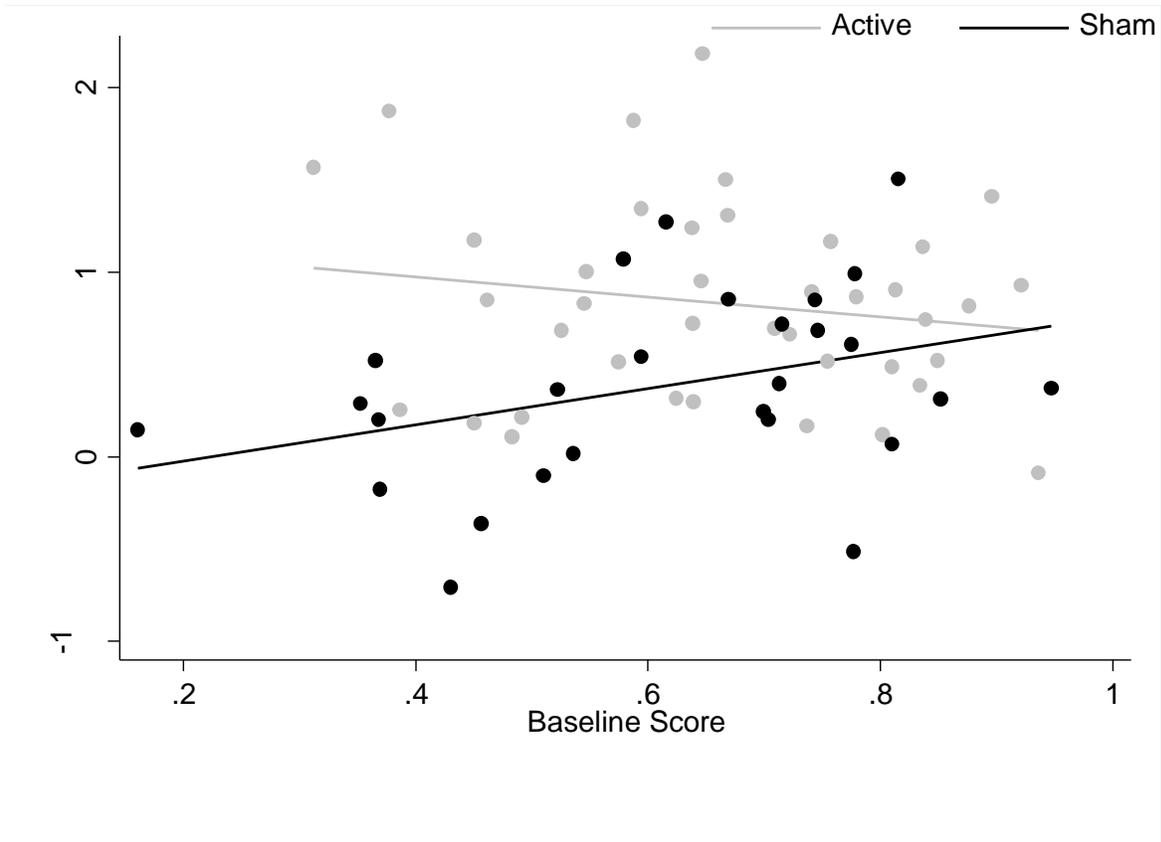
Active conditions. A standard linear regression was performed between training slope as the outcome variable and condition (Active and Sham) as the predictor variable. The condition factor was found to explain a significant amount of variance in the slope,  $F(1,65) = 11.65$ ,  $p < .001$ ,  $R^2 = .15$ ,  $R^2$  adjusted = .14. Condition significantly predicted slope (Beta = .79,  $t(66) = 3.41$ ,  $p = .001$ ) in that Active participants on average performed .79 standard deviations above Sham participants.

**Individual-difference Factors.** For each individual-difference factor, standard multiple regressions were performed between training slope as the outcome variable and condition, the individual difference, and the interaction between condition and the difference as predictor variables. Regression results are presented in Table 5.1.

<i>Model</i>	<i>Variable</i>	<i>N</i>	<i>B</i>	<i>SE B</i>	$\beta$	<i>p</i>	<i>Adj. R<sup>2</sup></i>
<i>Baseline</i>	Condition		1.41	0.34	0.76	.002	0.17
	Baseline WM	67	0.98	0.49	0.30	.07	
	Cond x baseline		-1.07	0.53	-0.47	.04	
<i>Motivation</i>	Condition		-1.63	0.68	0.81	<.001	0.25
	Motivation	67	-0.31	0.09	-0.64	.001	
	Cond x mot		0.35	0.11	0.73	.002	
<i>Gender</i>	Condition		0.45	0.16	0.78	.004	0.15
	Gender	67	0.25	0.21	0.44	.25	
	Cond x gender		-0.04	0.27	-0.06	.90	
<i>Site</i>	Condition		0.59	0.20	1.04	.004	0.13
	Training site	67	0.08	0.21	0.15	.69	
	Cond x site		-0.27	0.27	-0.48	.31	

Table 5.1. *Regression Results for Individual-difference measures.* Dummy coding of the categorical variables condition, gender, and training site employed the following references, respectively: sham, female, UCI. Unstandardized coefficients (**B**) are not mean-centered while standardized coefficients inherently are, and should be interpreted accordingly. For example, in the motivation model, **B** suggests a Sham advantage of 1.63 in the training slope when motivation is zero, while  $\beta$  suggests a tDCS advantage of .81 standard deviations when motivation is average.

**Baseline performance.** The model containing condition, baseline *n*-back performance, and the interaction term between condition and baseline performance explained a significant amount of variance in the training slope,  $F(3,63) = 5.53$ ,  $p = .002$ ,  $R^2 = .21$ ,  $R^2$  adjusted = .17. The partial effect of condition was significant (Beta = .76,  $t(66) = 3.30$ ,  $p = .002$ ) with larger slopes in the Active condition compared to Sham, holding baseline constant at the sample mean (i.e., baseline is mean-centered to zero). The partial effect of baseline, referenced to the Sham condition, suggests at the trend level that greater baseline performance is associated with larger slopes in the *absence* of tDCS (Beta = .30,  $t(66) = 1.83$ ,  $p = .07$ ). Importantly, the interaction term between condition and visual *n*-back performance at baseline was significant (Beta = -.47,  $t(66) = -2.06$ ,  $p = .04$ ), indicating that each standard deviation increase in baseline performance reduces the effect of condition by .47 standard deviations. A Bayesian follow-up analysis largely supported this result with  $BF_{10} = 17.07$  for the model with the baseline measure, condition, and the interaction term, and with  $BF_{10} = 5.63$  for the interaction term itself. This suggests that tDCS is most effective among low-baseline individuals (Figure 5.1).



*Figure 5.1.* Plot of baseline regression results. Active participants with low baseline scores outperform sham participants with low baseline scores, but the tDCS advantage gradually disappears with increasing baseline ability. Individuals with high baseline ability all improve similarly on the training task, regardless of condition.

**Motivation.** The model containing condition, motivation, and the interaction term between condition and motivation also explained a significant amount of the variance in the training slope  $F(3,63) = 8.45, p < .001$ , with an  $R^2 = .29, R^2 \text{ adjusted} = .25$ . The partial effect of condition, holding motivation constant at the mean, was significant (Beta = .81,  $t(66) = 3.78, p < .001$ ), reiterating the superior performance of the Active condition. However, the partial effect of motivation referenced to the Sham condition was also significant (Beta = -.64,  $t(66) = -3.41, p = .001$ ), as was the interaction term between condition and motivation (Beta = .73,  $t(66) = 3.15, p = .002$ ), suggesting somewhat paradoxically that within the sham group, participants with self-reported

higher motivation perform worse than participants with lower motivation (Figure 5.2). A Bayesian follow-up analysis supported this primary result with  $BF_{10} = 302.61$  for the model with the motivation average, condition, and the interaction term, and with  $BF_{10} = 26.96$  for the interaction term itself.

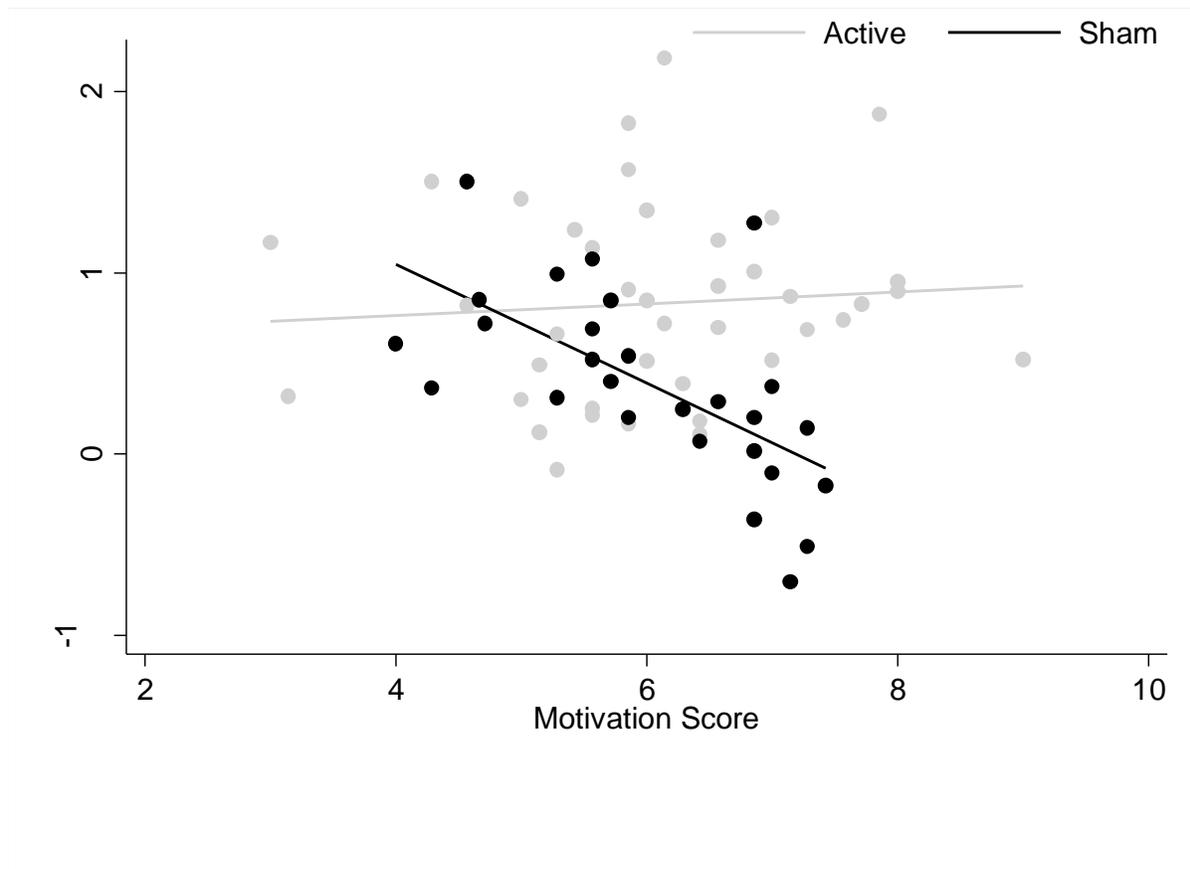


Figure 5.2. Plot of motivation regression results. Active participants all improve similarly irrespective of motivation, but sham participants show a paradoxical decrease in performance with increasing motivation.

**Gender.** The model containing condition, gender, and the interaction term between condition and gender explained a significant amount of the variance in the training slope  $F(3,63) = 4.93, p = .004$ , with an  $R^2 = .19, R^2$  adjusted = .15. While the partial effect of condition holding

gender constant among women was significant (Beta = .78,  $t(66) = 2.73$ ,  $p < .01$ ), neither gender (Beta = .44,  $t(66) = 1.17$ ,  $p = .25$ ), nor the interaction term between condition and gender (Beta = -.06,  $t(66) = -.13$ ,  $p = .90$ ) was significant. However, a Bayesian follow-up analysis resulted in a  $BF_{10} = 9.26$  for the model with the gender, condition, and the interaction term, and with  $BF_{10} = 4.15$  for the interaction term itself. Thus, there is likely value in continuing to examine the impact of gender in tDCS work.

**Study Site.** The model containing condition, site of training (i.e. UM or UCI), and the interaction term between condition and site also explained a significant amount of the variance in the slope  $F(3, 63) = 4.33$ ,  $p = .008$ , with an  $R^2 = .17$ ,  $R^2$  adjusted = .13. Again, while condition was a significant predictor (Beta = 1.04,  $t(66) = 2.98$ ,  $p = .004$ ), neither training site (Beta = .15,  $t(66) = .40$ ,  $p = .69$ ), nor the interaction term between condition and training site (Beta = .48,  $t(66) = 1.02$ ,  $p = .31$ ) was significant. A Bayesian follow-up analysis generally supported this primary result with  $BF_{10} = 4.94$  for the model with the gender, condition, and the interaction term, and with  $BF_{10} = 0.33$  for the interaction term itself.

**Long-term Follow-up.** An ANCOVA was conducted with condition as a factor, time between the intervention and the follow-up as a covariate, and gain on the training task from the first training session to the second follow-up as the dependent variable to evaluate whether an effect of condition remained at the second follow-up that took place, on average, 355 days following the conclusion of the intervention. Condition remained a significant factor for the second follow-up,  $F(1,23) = 12.43$ ,  $p = .002$ , with Active participants continuing to outperform Sham participants (Figure 5.3), while, as in the first follow-up reported in Au et al. (2016), time between the intervention and follow-up was not a significant predictor  $F(1,23) = 1.18$ ,  $p = .29$ .

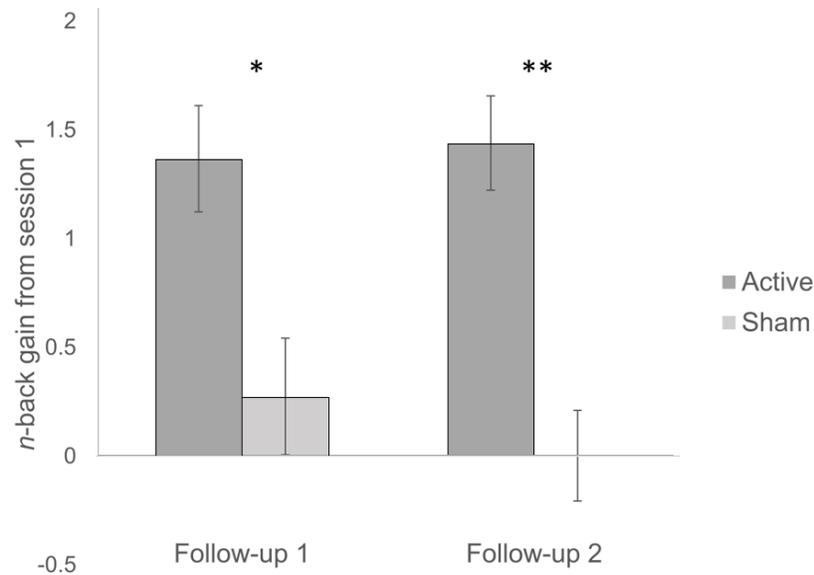


Figure 5.3. Follow-up performance. Follow-up 1 represents in n-back levels gain from the first session to the first follow-up for active and sham participants reported in Au et al. (2016); Follow-up 2 represents gain from the first session to the new second follow-up approximately 12 months after the intervention

### Discussion

Here we present evidence that certain individual-difference factors do have a significant impact on the outcome of combined WM training and tDCS. The effect of baseline was particularly striking. We found a trend suggesting that Sham participants who started with higher baseline ability tended to improve more over the course of training. Though this finding did not reach traditional levels of statistical significance ( $p=.07$ ), it is nevertheless consistent with previous literature suggesting cognitive training may be more helpful to those who already have strong cognitive abilities (Looi et al., 2016; Lövdén, Brehmer, Li, & Lindenberger, 2012). More importantly, however, the interaction between baseline ability and condition (Active/Sham) was significant (see Figure 5.1), suggesting that the effects of baseline ability affected Active and

Sham participants differently. Specifically, the advantage of tDCS seemed to increase proportionately with decreasing baseline ability, such that a participant who started off 1 standard deviation below the mean in terms of visual WM ability before training ended up outperforming a comparable Sham participant by .46 standard deviations over the course of training. However, this tDCS advantage declines with increasing baseline ability, and confers little additional advantage to a participant who already performs high at baseline relative to a comparably high-performing peer in the Sham group. Although it is unclear what may mediate this interaction between stimulation and low baseline performance, it may have to do with differences in brain state and baseline cortical excitability between high and low groups (c.f., Krause & Cohen Kadosh, 2014). For example, it is known that the effects of transcranial magnetic stimulation (TMS) are influenced by the baseline excitability of the targeted cortex (Pasley, Allen, & Freeman, 2009; Silvanto, Cattaneo, Battelli, & Pascual-Leone, 2008) and that lower or more suppressed levels of neural excitability can increase the facilitatory effect of TMS. That said, these findings should be tempered by the fact that these findings would likely not survive multiple comparison tests; even if a fairly liberal Bonferroni correction accounting for only the primary tests of the interaction term was used (thus requiring a  $p$  value below .0125), the baseline result would not remain significant, although the curious motivation result would still stand.

We note that this finding of selective tDCS-enhancement among low-baseline individuals is not unique in the literature. For example, a number of studies also suggest a selective tDCS benefit among low-baseline populations, both within the WM domain (Gozenman & Berryhill, 2016; Heinen et al., 2016; Minichino et al., 2015; Tseng et al., 2012) as well as in other cognitive domains, such as attention and dual-tasking (London & Slagter, 2015; Reinhart et al., 2016;

Zhou et al., 2015). However, one critical difference between these studies and ours is that ours is a training study involving multiple sessions of stimulation in conjunction with task performance, rather than just a single session (but see also Looi et al., 2016). Consequently, these results demonstrate enhancements not only to overall WM performance, but more specifically, to the rate of learning (as measured by the slope of improvement) across sessions. This raises the possibility that the selective effects of stimulation on low-baseline participants may impact not only online performance, but also offline consolidation, an important distinction for the enhancement of long-term learning (Au, Karsten, Buschkuehl, & Jaeggi, in press). Though these offline effects were supported in our previous work by demonstrating especial tDCS benefits when training sessions were spaced apart by a weekend (Au et al., 2016), a possible interaction of baseline performance and weekend consolidation in the present work is difficult to demonstrate due to power issues. For the same reason, the influence of baseline ability on follow-up performance is similarly difficult to evaluate.

While self-reported motivation also had a significant impact on the outcome of training, the finding of a significant interaction between motivation and condition was somewhat puzzling. The nature of the interaction is such that motivation is inversely related to slope in the Sham group only. It is unclear why lower-motivated individuals outperformed higher-motivated individuals in the Sham condition, but one possibility is that lower motivation was also associated with other influential factors, such as higher baseline performance (it is possible that for individuals who performed very well already, the intervention was not as interesting, while those who were aware of pre-existing limitations were more eager to improve their cognitive abilities). In fact, there is a moderately strong inverse correlation between baseline and motivation within the Sham group ( $r=-.42$ ), suggesting that some of the observed motivation

effect simply recapitulates the baseline effect. Nevertheless, we also note that both high and low motivated individuals within the Active group experienced similar improvement during the intervention, suggesting that for those individuals receiving stimulation, motivation was not a major factor that impacted performance. We also note that our motivation measure—a single question asked each day before training—may be less ideal than a more in-depth survey measure (and such a measure might be better equipped to explain the curious motivation results discussed here). Finally, neither gender, nor site of training, nor the interaction between those variables and condition, predicted the slope of training. Thus, these analyses provide evidence that some individual-difference factors, such as baseline WM performance, play a major role in the outcome of combined tDCS and cognitive training, while others do not.

Within the context of the larger corpus of tDCS research, these findings have significant implications for both existing and future studies that combine cognitive training with stimulation. Given the extent to which these factors, including baseline performance in particular, influence the outcome of training, it is possible that these differences may explain why so many participants in any individual study do not benefit from stimulation. Furthermore, it may also explain some of the null findings and even some of the varied outcomes observed among successful studies. At the very least, these findings provide an impetus for examining baseline differences as a covariate of interest in training and stimulation studies. This also means that future studies must be adequately powered to account for these differences and allow for them to be examined.

We also note the continued difference between the Active and Sham conditions approximately a year (on average) following the intervention, extending the medium-term follow-up findings established in Au et al. (2016). This suggests that applying tDCS with

cognitive training may not only result in more robust and rapid improvements on the training task, but also that the improved performance on the training task relative to the Sham group may remain stable, even up to a year after the intervention. Importantly, we note that this follow-up examined only training effects, rather than any improvements in transfer tasks. Future work will be needed to establish the extent that transfer gain may also persist at long-term follow-up.

We note that these results must be tempered by certain limitations in our dataset. The baseline measure included here is perhaps less ideal than having the participant complete a session of the training task prior to stimulation, which would give a “true baseline” that might be a better predictor of subsequent performance. Additionally, there was considerable attrition between the initial study and the second follow-up. Finally, although this study was fairly well-powered for a tDCS and training intervention, even more participants would be needed to have better confidence about the individual-difference findings presented here. Furthermore, we acknowledge that this study is not powered well enough to examine more than one individual-difference factor at a time, and the follow-up sample is too small to examine in the context of the individual-difference factors covered here. Thus it is important to note the preliminary nature of the present analyses.

Despite these limitations, the practical implications of the baseline finding are of particular interest, both for cognitive training studies as well as tDCS-augmented learning more generally. Within cognitive training research, some studies have suggested that it is necessary for participants to demonstrate improvement on the training task in order to achieve transfer gains (e.g., Jaeggi et al., 2011). TDCS may enable participants with lower starting performance to reach similar gains to their higher-performing peers, thus overriding individual differences in baseline ability, and allowing more to benefit from the intervention. In the context of learning

more generally, tDCS may offer a means of helping individuals who might be struggling on a particularly WM-demanding task, such as math, improve more quickly. Preliminary research, albeit with only two sessions, suggests that this may indeed be possible (Looi et al., 2016). Additionally, subsequent work should combine this line of investigation with fMRI or EEG; the combination of physiological or neuroimaging data may allow researchers to better understand how physical characteristics and anatomical differences may impact the flow of current generated by the stimulation device. Most importantly, these results reinforce the importance of considering individual-differences during the administration of tDCS and training—as well as collecting samples well-powered enough to actually examine them.

## References

- Alloway, T. P., & Alloway, R. G. (2010). Investigating the predictive roles of working memory and IQ in academic attainment. *Journal of Experimental Child Psychology, 106*, 20-29.
- Au, J., Karsten, C., Buschkuehl, M., & Jaeggi, S. M. (in press). Optimizing Transcranial Direct Current Stimulation Protocols to Promote Long-Term Learning. *Journal of Cognitive Enhancement*.
- Au, J., Katz, B., Buschkuehl, M., Bunarjo, K., Senger, T., Zabel, C., et al. (2016). Enhancing Working Memory Training with Transcranial Direct Current Stimulation. *Journal of Cognitive Neuroscience, 28*(9), 1419-1432.
- Au, J., Sheehan, E., Tsai, N., Duncan, G. J., Buschkuehl, M., & Jaeggi, S. M. (2015). Improving fluid intelligence with training on working memory: a meta-analysis. *Psychon Bull Rev, 22*(2), 366-377.
- Benwell, C. S., Learmonth, G., Miniussi, C., Harvey, M., & Thut, G. (2015). Non-linear effects of transcranial direct current stimulation as a function of individual baseline performance: Evidence from biparietal tDCS influence on lateralized attention bias. *Cortex, 69*, 152-165.
- Berryhill, M. E., & Jones, K. T. (2012). tDCS selectively improves working memory in older adults with more education. *Neuroscience Letters, 521*(2), 148-151.
- Borella, E., Carretti, B., Cantarella, A., Riboldi, F., Zavagnin, M., & De Beni, R. (2014). Benefits of Training Visuospatial Working Memory in Young-Old and Old-Old. *Developmental Psychology, 50*(3), 714-727.
- 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*, 63.
- Brunoni, A. R., Kemp, A. H., Shiozawa, P., Cordeiro, Q., Valiengo, L. C. L., Goulart, A. C., et al. (2013). Impact of 5-HTTLPR and BDNF polymorphisms on response to sertraline versus transcranial direct current stimulation: Implications for the serotonergic system. *European Neuropsychopharmacology, 23*(11), 1530-1540.
- Buzsáki, G. (Ed.) (2016) Cognitive Neuroscience Society 23rd Annual Meeting. New York, New York.
- Dedoncker, J., Brunoni, A. R., Baeken, C., & Vanderhasselt, M. A. (2016). A Systematic Review and Meta-Analysis of the Effects of Transcranial Direct Current Stimulation (tDCS) Over the Dorsolateral Prefrontal Cortex in Healthy and Neuropsychiatric Samples: Influence of Stimulation Parameters. *Brain Stimulation, 9*(4), 501-517.
- Gozenman, F., & Berryhill, M. E. (2016). Working memory capacity differentially influences responses to tDCS and HD-tDCS in a retro-cue task. *Neuroscience Letters, 629*, 105-109.

- Heinen, K., Sagliano, L., Candini, M., Husain, M., Cappelletti, M., & Zokaei, N. (2016). Cathodal transcranial direct current stimulation over posterior parietal cortex enhances distinct aspects of visual working memory. *Neuropsychologia*, *87*, 35-42.
- Higgins, D. M., Peterson, J. B., Pihl, R. O., & M, G. (2007). Prefrontal cognitive ability, intelligence, Big Five personality, and the prediction of advanced academic and workplace performance. *Journal of Personality and Social Psychology*, *93*, 298-319.
- Hill, A. T., Fitzgerald, P. B., & Hoy, K. E. (2016). Effects of Anodal Transcranial Direct Current Stimulation on Working Memory: A Systematic Review and Meta-Analysis of Findings From Healthy and Neuropsychiatric Populations. *Brain Stimulation*, *9*(2), 197-208.
- Hsu, W. Y., Ku, Y., Zanto, T. P., & Gazzaley, A. (2015). Effects of noninvasive brain stimulation on cognitive function in healthy aging and Alzheimer's disease: a systematic review and meta-analysis. *Neurobiology of Aging*, *36*(8), 2348-2359.
- Jaeggi, S. M., Buschkuhl, M., Jonides, J., & Shah, P. (2011). Short- and long-term benefits of cognitive training. *Proceedings of the National Academy of Sciences of the USA*, *108*(25), 10081-10086.
- Jaeggi, S. M., Buschkuhl, M., Shah, P., & Jonides, J. (2014). The role of individual differences in cognitive training and transfer. *Memory & Cognition*, *42*(3), 464-480.
- Jeon, S. Y., & Han, S. J. (2012). Improvement of the working memory and naming by transcranial direct current stimulation. *Annals of Rehabilitative Medicine*, *36*(5), 585-595.
- Jones, K. T., & Berryhill, M. E. (2012). Parietal contributions to visual working memory depend on task difficulty. *Frontiers in Psychiatry*, *3*, 81.
- Jones, K. T., Gozenman, F., & Berryhill, M. E. (2015). The strategy and motivational influences on the beneficial effect of neurostimulation: a tDCS and fNIRS study. *Neuroimage*, *105*, 238-247.
- Jones, K. T., Stephens, J. A., Alam, M., Bikson, M., & Berryhill, M. E. (2015). Longitudinal Neurostimulation in Older Adults Improves Working Memory. *Plos One*, *10*(4).
- Karbach, J., & Verhaeghen, P. (2014). Making working memory work: a meta-analysis of executive-control and working memory training in older adults. *Psychological Science*, *25*(11), 2027-2037.
- Katz, B., Jaeggi, S., Buschkuhl, M., Stegman, A., & Shah, P. (2014). Differential effect of motivational features on training improvements in school-based cognitive training. *Frontiers in Human Neuroscience*, *8*, 242.
- Katz, B., Jones, M. R., Shah, P., Buschkuhl, M., & Jaeggi, S. M. (in press). Individual difference and motivational effects in cognitive training research. In T. Stroback & J. Karbach (Eds.), *Cognitive Training: An Overview of Features and Applications*. Berlin: Springer.

- Kim, J. H., Kim, D. W., Chang, W. H., Kim, Y. H., Kim, K., & Im, C. H. (2014). Inconsistent outcomes of transcranial direct current stimulation may originate from anatomical differences among individuals: electric field simulation using individual MRI data. *Neuroscience Letters*, *564*, 6-10.
- Klingberg, T., Fernell, E., Olesen, P. J., Johnson, M., Gustafsson, P., Dahlstrom, K., et al. (2005). Computerized training of working memory in children with ADHD--a randomized, controlled trial. *Journal of the American Academy of Child & Adolescent Psychiatry*, *44*(2), 177-186.
- Krause, B., & Cohen Kadosh, R. (2014). Not all brains are created equal: the relevance of individual differences in responsiveness to transcranial electrical stimulation. *Frontiers in Systems Neuroscience*, *8*, 25.
- Krause, B., Marquez-Ruiz, J., & Cohen Kadosh, R. (2013). The effect of transcranial direct current stimulation: a role for cortical excitation/inhibition balance? *Frontiers in Human Neuroscience*, *7*, 602.
- Learmonth, G., Thut, G., Benwell, C. S., & Harvey, M. (2015). The implications of state-dependent tDCS effects in aging: Behavioural response is determined by baseline performance. *Neuropsychologia*, *74*, 108-119.
- London, R. E., & Slagter, H. A. (2015). Effects of Transcranial Direct Current Stimulation over Left Dorsolateral pFC on the Attentional Blink Depend on Individual Baseline Performance. *Journal of Cognitive Neuroscience*, *27*(12), 2382-2393.
- Looi, C. Y., Duta, M., Brem, A. K., Huber, S., Nuerk, H. C., & Kadosh, R. C. (2016). Combining brain stimulation and video game to promote long-term transfer of learning and cognitive enhancement. *Scientific Reports*, *6*.
- López-Alonso, V., Cheeran, B., Río-Rodríguez, D., & Fernández-del-Olmo, M. (2014). Inter-individual Variability in Response to Non-invasive Brain Stimulation Paradigms. *Brain Stimulation*, *7*, 372-380.
- Lövdén, M., Bäckman, L., Lindenberger, U., Schaefer, S., & Schmiedek, F. (2010). A theoretical framework for the study of adult cognitive plasticity. *Psychological Bulletin*, *136*, 659-676.
- Lövdén, M., Brehmer, Y., Li, S.-C., & Lindenberger, U. (2012). Training-induced compensation versus magnification of individual differences in memory performance. *Frontiers in Human Neuroscience*, *6*, 141.
- Mancuso, L. E., Ilieva, I. P., Hamilton, R. H., & Farah, M. J. (2016). Does Transcranial Direct Current Stimulation Improve Healthy Working Memory?: A Meta-analytic Review. *Journal of Cognitive Neuroscience*, *28*, 1063-1089.

- Manenti, R., Sandrini, M., Brambilla, M., & Cotelli, M. (2016). The optimal timing of stimulation to induce long-lasting positive effects on episodic memory in physiological aging. *Behavioral Brain Research*, *311*, 81-86.
- Martin, D. M., Liu, R., Alonzo, A., Green, M., Player, M. J., Sachdev, P., et al. (2013). Can transcranial direct current stimulation enhance outcomes from cognitive training? A randomized controlled trial in healthy participants. *International Journal Neuropsychopharmacol*, *16*(9), 1927-1936.
- McCambridge, A. B., Bradnam, L. V., Stinear, C. M., & Byblow, W. D. (2011). Cathodal transcranial direct current stimulation of the primary motor cortex improves selective muscle activation in the ipsilateral arm. *Journal of Neurophysiology*, *105*(6), 2937-2942.
- Melby-Lervåg, M., & Hulme, C. (2013). Is working memory training effective? A meta-analytic review. *Developmental Psychology*, *49*(2), 270-291.
- Minichino, A., Bersani, F. S., Bernabei, L., Spagnoli, F., Vergnani, L., Corrado, A., et al. (2015). Prefronto-cerebellar transcranial direct current stimulation improves visuospatial memory, executive functions, and neurological soft signs in patients with euthymic bipolar disorder. *Neuropsychiatric Disease and Treatment*, *11*, 2265-2270.
- Park, D. C., Lautenschlager, G., Hedden, T., Davidson, N. S., Smith, A. D., & Smith, P. K. (2002). Models of visuospatial and verbal memory across the adult life span. *Psychology and Aging*, *17*(2), 299-320.
- Park, S. H., Seo, J. H., Kim, Y. H., & Ko, M. H. (2014). Long-term effects of transcranial direct current stimulation combined with computer-assisted cognitive training in healthy older adults. *Neuroreport*, *25*(2), 122-126.
- Pasley, B. N., Allen, E. A., & Freeman, R. D. (2009). State-dependent variability of neuronal responses to transcranial magnetic stimulation of the visual cortex. *Neuron*, *62*(2), 291-303.
- Plewnia, C., Zwissler, B., Langst, I., Maurer, B., Giel, K., & Kruger, R. (2013). Effects of transcranial direct current stimulation (tDCS) on executive functions: Influence of COMT Val/Met polymorphism. *Cortex*, *49*(7), 1801-1807.
- Prins, P. J. M., DAVIS, S., Ponsioen, A., ten Brink, E., & van der Oord, S. (2010). Does Computerized Working Memory Training with Game Elements Enhance Motivation and Training Efficacy in Children with ADHD? *Cyberpsychology, Behavior, and Social Networking*, *14*, 115-122.
- Redick, T. S., & Lindsey, D. R. (2013). Complex span and n-back measures of working memory: a meta-analysis. *Psychonomic Bulletin and Review*, *20*(6), 1102-1113.
- Reinhart, R. M., Xiao, W., McClenahan, L. J., & Woodman, G. F. (2016). Electrical Stimulation of Visual Cortex Can Immediately Improve Spatial Vision. *Current Biology*, *26*(14), 1867-1872.

- Reis, J., Schambra, H. M., Cohen, L. G., Buch, E. R., Fritsch, B., Zarahn, E., et al. (2009). Noninvasive cortical stimulation enhances motor skill acquisition over multiple days through an effect on consolidation. *Proceedings of the National Academy of Sciences of the USA*, *106*(5), 1590-1595.
- Schmiedek, F., Lovden, M., & Lindenberger, U. (2010). Hundred Days of Cognitive Training Enhance Broad Cognitive Abilities in Adulthood: Findings from the COGITO Study. *Front Aging Neurosci*, *2*.
- Segretin, M. S., Lipina, S. J., Hermida, M. J., Sheffield, T. D., Nelson, J. M., Espy, K. A., et al. (2014). Predictors of cognitive enhancement after training in preschoolers from diverse socioeconomic backgrounds. *Frontiers in Psychology*, *5*.
- Sikstrom, S., Jurgensen, A. M., Haghghi, M., Mansson, D., Smidelik, D., & Habekost, T. (2016). Self-Rated Attentiveness Interacts with Transcranial Direct Current Stimulation and Noise Stimulation in Reaction Time in a Go/No-Go Task. *Neural Plasticity*, *2016*, 5302538.
- Silvanto, J., Cattaneo, Z., Battelli, L., & Pascual-Leone, A. (2008). Baseline cortical excitability determines whether TMS disrupts or facilitates behavior. *Journal of Neurophysiol*, *99*(5), 2725-2730.
- Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: applications to dementia and amnesia. *Journal of Experimental Psychology General*, *117*(1), 34-50.
- Söderqvist, S., Bergman Nutley, S., Ottersen, J., Grill, K. M., & Klingberg, T. (2012). Computerized training of non-verbal reasoning and working memory in children with intellectual disability. *Frontiers in Human Neuroscience*, *6*, 271.
- Studer-Luethi, B., Bauer, C., & Perrig, W. J. (2015). Working memory training in children: Effectiveness depends on temperament. *Memory & Cognition*, *44*, 171-186.
- Studer-Luethi, B., Jaeggi, S. M., Buschkuhl, M., & Perrig, W. J. (2012). Influence of neuroticism and conscientiousness on working memory training outcome. *Personality and Individual Differences*, *53*(1), 44-49.
- Summers, J. J., Kang, N., & Cauraugh, J. H. (2015). Does transcranial direct current stimulation enhance cognitive and motor functions in the ageing brain? A systematic review and meta-analysis. *Ageing Res Rev*, *25*, 42-54.
- Tseng, P., Hsu, T. Y., Chang, C. F., Tzeng, O. J. L., Hung, D. L., Muggleton, N. G., et al. (2012). Unleashing Potential: Transcranial Direct Current Stimulation over the Right Posterior Parietal Cortex Improves Change Detection in Low-Performing Individuals. *The Journal of Neuroscience*, *32*, 10554-10561.

- Uehara, K., Coxon, J. P., & Byblow, W. D. (2015). Transcranial direct current stimulation improves ipsilateral selective muscle activation in a frequency dependent manner. *PLoS One*, *10*(3), e0122434.
- Wass, S. V., Scerif, G., & Johnson, M. H. (2012). Training attentional control and working memory - Is younger, better? *Developmental Review*, *32*(4), 360-387.
- Wiethoff, S., Hamada, M., & Rothwell, J. C. (2014). Variability in Response to Transcranial Direct Current Stimulation of the Motor Cortex. *Brain Stimulation*, *7*, 468-475.
- Zhou, D., Zhou, J., Chen, H., Manor, B., Lin, J., & Zhang, J. (2015). Effects of transcranial direct current stimulation (tDCS) on multiscale complexity of dual-task postural control in older adults. *Experimental Brain Research*, *233*(8), 2401-2409.
- Zinke, K., Zeintl, M., Eschen, A., Herzog, C., & Kliegel, M. (2012). Potentials and Limits of Plasticity Induced by Working Memory Training in Old-Old Age. *Gerontology*, *58*, 79-87.
- Zinke, K., Zeintl, M., Rose, N. S., Putzmann, J., Pydde, A., & Kliegel, M. (2014). Working memory training and transfer in older adults: Effects of age, baseline performance, and training gains. *Developmental Psychology*, *50*, 304-315.

## CHAPTER 6

### The Map is Not the Territory

#### Conclusion

The aim of this dissertation, from the most general perspective, has been to better understand why some cognitive training studies are successful and others are not. More specifically, I investigated one particular individual-difference factor, motivation, in three working memory training studies. Across these three studies, I found that motivational factors—including game-like engagement features, recruitment posters and compensation, and self-ratings of motivation prior to training—sometimes impact the outcome of training. However, this impact was inconsistent and often idiosyncratic. Inconsistent results, however, are extremely common in cognitive training research, and, as I will discuss in this conclusion, may be indicative of larger issues facing the field.

#### Summary of Individual Studies

Study 1 explored whether game elements that are often considered motivational—such as scoring or leveling—improved the outcome of a brief training intervention. Contrary to our hypothesis, the versions of the game that contained the *fewest* motivational features actually led to the greatest training and near-transfer gains. Study 2 examined whether monetary compensation—a reward that would likely be defined as “extrinsic” motivation—led to reduced training or transfer performance versus training conducted with no external reward. In fact, the

compensation provided had no effect on the outcome of the study. Although there were some differences in baseline characteristics between the two training groups, both experienced significant (and roughly equivalent) gains versus an active control group. Finally, Study 3 examined a number of individual-difference factors, including motivation, in the context of tDCS-augmented cognitive training. Interestingly, the individual-difference factors that seemed to impact the outcome of training for the group that received sham stimulation (baseline and motivation) had no impact on the performance of the stimulated condition. However, like Studies 1 and 2, the relationship between motivation and performance was surprising (at least for the sham condition): those participants with *lower* motivation scores actually performed better at the training. It is worth noting, as well, that Studies 1 and 3 focused primarily at training performance, while Study 2 also focused on far transfer (Study 1 only included a near-transfer measure). While training performance and transfer are often closely connected, this is not always the case, and the possible interplay between motivation, training, *and* transfer together is likely worthy of future investigation.

Our hypotheses were not fully supported, to some extent, in each of the three studies. What can be made of these unexpected results? It remains possible that some other factor that was not measured is at play; it is also possible, as with most cognitive training studies, that low-power led to these surprising results. It is also possible, however, that motivation is a complex construct and that its impact on the outcome of extended interventions may often be more nuanced than expected. For example, in Study 1, it is possible that “motivational elements” such as those included in games might actually be beneficial over the long term, even if they are apparently harmful to early learning of a task. In Study 2, it remains possible that compensation could have an undermining effect on motivation and performance in certain situations. However,

the manipulation check used in the paid group provides some preliminary evidence that the motivational orientation of our participants defies “intrinsic” versus “extrinsic” divisions. Many participants reported being self-motivated to participate, despite seeing no self-improvement language. This was an unexpected result. Finally, in Study 3, stimulation *did* have the effect of seemingly minimizing the effects of the examined individual differences on the outcome of the training. That said, how can we explain the relationship between motivation and training performance in the sham group? Why would more highly-motivated participants perform worse? One possible answer to this question is that a single motivation question is not adequate to fully explore the impact of motivation on the outcome of training. This would certainly come as no surprise to motivation researchers. In aggregate, the present findings tell us more about cognitive training (or tDCS) more generally than motivation in particular. Study 1 and Study 2 provide modest support for the existence of near transfer following a training intervention; while Study 3 suggests that tDCS may reduce the impact of certain individual difference factors on training performance. Without larger sample sizes, longer training regimens, and a better understanding of how individuals who dropped out would respond if they had completed the training, it is difficult to draw strong, or even modest, conclusions about how the motivational factors of interest would impact individuals in a real-world intervention. In hindsight, it is easy to identify these limitations; it is worth noting that some of them were considered even before running these studies. These examples demonstrate how challenging running laboratory-based cognitive training studies can be. Thus, while we have arguably provided some preliminary evidence that motivation may play a nuanced role in the outcome of cognitive training, it is also clear that this relationship may require considerable research. Of course, this is a common refrain at the

conclusion of dissertations, but perhaps there are other lessons that can be learned from this line of work.

### **The Problem with Binary Questions**

It is likely accurate to state that the question most-often asked of cognitive training researchers is whether cognitive training “works.” Perhaps this is the wrong question to ask. Alan Newell, in his classic piece that inspired our title (Newell, 1973), points out that psychological science operates on two levels: one in which there are incremental, specific studies that elaborate on a phenomenon (e.g., does cognitive training work better when practice is spaced or massed) and one that asks fairly large binary questions (e.g., nature vs. nurture). But what is missed is a more unified approach that allows us to better understand “the behavior of man” (p. 6). Newell offers one highly relevant strategy to achieve this: to center experimental and theoretical work around “a single complex task,” and in the service of this develop a coherent theoretical model supported by many smaller studies (e.g., Meyer & Kieras, 1997). If cognitive training researchers were to take up his recommendations, they would need to develop computational process models of prefrontal function and intelligent behavior. Hypotheses about how the model improved on the training task and how that would generalize to a transfer task could then be tested via the model. Empirical studies would support model development via micro-studies that help generate parameters, and also studies that test the model’s predictions. In particular, empirical studies should test possible underlying mechanisms that support transfer. As the Buffalo Springfield lyric goes, “there’s something happening here, but what it is ain’t exactly clear.” In cognitive training we have identified what appears to be a compelling phenomenon,

but without an overarching theoretical framework to guide empirical research, progress in understanding this phenomenon will likely be stalled.

Given this, one reasonable question is what the ideal “motivation in cognitive training” study might look like. Could one design a study (or, perhaps, a set of studies) to better answer the sort of questions addressed earlier? For example, let’s say we were specifically interested in verifying whether game-like elements played a role in the outcome of training, in addition to whether recruitment materials and compensation impacted outcomes, and finally, whether motivational orientation throughout the intervention impacted the outcome of training. This might seem like too many questions to address in a single experiment, but in fact, this sort of complexity is addressed through A/B testing within large internet corporations on a daily basis. Ideally, one would include both *control* (likely a trivia-based task) and *experimental* task versions of the game all the game-like features included, none of them, or a single feature removed. So instead of seven, fourteen groups would be required. Furthermore, each of these groups would also be recruited through either an advertisement that emphasized improving intelligence or an advertisement that emphasized simply completing some tasks for compensation, giving us twenty-eight groups total. Additionally, each of *these* groups would then be divided again, to determine whether the use of compensation impacted training and transfer. For the purposes of this “ideal study” we can also assume that five covariates might be included in the analysis (including, of course, full motivation surveys rather than single question, such as Need for Cognition, the Intrinsic Motivation Questionnaire, etc.). A power analysis with G\*Power, with these fifty-six groups and five covariates, and with a small effect size of .1 (similar to what was found in the studies in this dissertation) requires a fairly large number of participants: 2448. While this may seem like a daunting sample size, it is trivial compared to the

number of participants who complete online cognitive training each day through commercial brain training companies, and at least one study recruited far more participants using the internet (Owen et al., 2010). In this case, internet recruitment and online administration would be used as well, with different online ads used to target the “intrinsic” versus “extrinsic” participants.

Participants would complete a full twenty-sessions of training online and would also complete a large battery of pre- and post-tests as in Study 2 (and these tests could likely be reduced to a smaller number of latent or composite variables). Naturally, some participants would drop-out at different points of the study, and one would use the pre-test measures collected to determine what, if any differences existed between the drop-out and completed participants (although another, separate study would likely be used to fully explore the drop-out issue). At the end of the study, a researcher would then be able to examine not only the main effects of these different factors, but also interactions between them, looking not only at training performance but also transfer. Multiple comparisons would be an issue, but remember that even this large study would only be the first of several in the pursuit of developing a coherent model of motivation in cognitive training. Groups could be adjusted and new comparisons added based on the outcome of the study – for example, if removing score positively impacted performance but nothing else had this effect, the following study would simply include a version of training without score included during play, and a new comparison could be examined.

### **Implications for the Field**

As discussed earlier, there are often practical limits on conducting high quality studies of cognitive training. While it is valuable to work towards developing better studies of cognitive training, there remain other conflicts of interest and motivational factors that influence which

studies are conducted and ultimately published (90). In general, the studies that are most likely to appear in the press or high-impact journals are those that have novel, unexpected, and clearly impactful results. Studies with null effects, or those that replicate and incrementally test the boundary conditions of a finding, are perceived as much less valuable. Additionally, scientists are not immune to the idea of “motivated reasoning.” If they have strong beliefs or motivations inconsistent with the results of the study, they are easily able to find flaws (see the classic study by Lord, Ross, & Lepper, 1979). But when they wish to believe a finding, the flaws are less visible.

If there were no multi-million-dollar cognitive training industry, the field would be much less controversial. And yet, this is the world we live in. So what do we tell parents who want to know whether these programs can help their children? We know that proper nutrition (Glewwe, Jacoby, & King, 2001), sleep (Curcio, Gerrara, & De Gennaro, 2006), and physical exercise (Tomporowski et al., 2007) are beneficial for cognitive development, and such factors need to be addressed whether or not children engage in cognitive training. However, it is clear that there is little to be lost, and possibly much to be gained, through engaging in cognitively enriching activities (e.g. cognitive training but also music, dance, meditation, board games, etc.). At the same time, we hope that consumers will be on guard against strong promises offered by purveyors of cognitive training programs. Consider the continuing allure of “brain-based” marketing techniques. While recent work has provided experimental evidence for these strategies as tools of persuasion (Weisberg et al, 2008; Rhodes, Rodriguez, & Shah, 2014), their dangerous efficacy has been clear since the days of Pelmanism over a century ago. When neuroscience is evoked, non-experts are more likely to believe explanations—even if those explanations are otherwise unsound.

If one thing is certain, it is that the public interest in improving cognition will continue for the foreseeable future. But the outcome of any individual study, of any individual intervention, and, as we have illustrated, any individual meta-analysis cannot be construed as a conclusive answer to the question of how much cognitive function might be improved through intervention. In addition to the theoretical and modeling work discussed above, we note that significant attention should be given to the careful communication of findings. Often the fault on this count lies not in conducting studies that have methodological limitations and potential alternative interpretations, as these studies might guide us towards better future work and a richer understanding of the phenomenon. Rather, the fault lies in interpreting the results of these limited studies as “proof” that cognitive training does or does not work. When studies are published in short-form journals or reported in press releases, claims tend to be exaggerated and the limitations receive short shrift. One study in the British Medical Journal recently analyzed 462 health science press releases and found that 40% of them overstate the implications of the findings (Sumner et al., 2014). Unfortunately, even when “hedging” language is present in scientific articles, most readers gloss over the details and focus on the main claim when reading about scientific studies (Norris, Phillips, & Korpan, 2003).

That the issues discussed in Chapter 1 remain in spite of the actual empirical work conducted, and possibly *because* of it as well, especially when researchers improperly generalize or ignore previous findings. And they were not solved despite decades of theoretical and methodological discussions that in many respects were not dissimilar from more recent reviews, such as Simons et al. (Simons et al., 2016). Thus the issue of whether cognitive training “works” was not settled in 1910, nor 1914, nor in all the years that followed. As Mead wrote in 1946: “A final question is this: how long will it take for the facts known about transfer to be used, and

adjustments to be made accordingly? One hundred years? Or never?" (Mead, 1946). We still do not have definitive answers to this question. Researchers may not have the answer 100 years hence. But if we keep asking, simply, "does cognitive training work," rather than investigating the mechanisms of transfer within a coherent theoretical framework, we will never have them at all.

## References

- Curcio G., Ferrara M., De Gennaro, L. (2006) Sleep loss, learning capacity and academic performance. *Sleep Medicine Reviews* 10(5), 323–337.
- Glewwe P., Jacoby H. G., King E. M. (2001) Early childhood nutrition and academic achievement: a longitudinal analysis. *Journal of Public Economics* 81(3), 345–368.
- Lord C. G., Ross L., Lepper M. R. (1979) Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology* 37(11), 2098–2109.
- Mead RA (1946) Transfer of training again. *Journal of Educational Psychology* 37(7), 391–397.
- Meyer D. E., Kieras D. E. (1997) A computational theory of executive cognitive processes and multiple-task performance: Part I. Basic mechanisms. *Psychological Review* 104(1), 3–65.
- Newell, A. (1973) You can't play 20 questions with nature and win: Projective comments on the papers of this symposium. Available at: <http://repository.cmu.edu/cgi/viewcontent.cgi?article=3032&context=compsci> [Accessed August 19, 2016].
- Norris, S. P., Phillips, L. M., & Korpan, C. A. (2003) University Students' Interpretation of Media Reports of Science and its Relationship to Background Knowledge, Interest, and Reading Difficulty. *Public Understanding of Science* 12(2), 123–145.
- Rhodes, R. E., Rodriguez, F., Shah, P. (2014) Explaining the alluring influence of neuroscience information on scientific reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 40(5), 1432–1440.
- Simons D. J., Boot, W. R., Charness, N., Gathercole, S. E., Chabris, C. F., Hambrick, D. Z., & Stine-Morrow, E. A. (2016) Do “Brain-Training” Programs Work? *Psychological Science in the Public Interest*, 17(3), 103-186.
- Sumner, P., Vivian-Griffiths, S., Boivin, J., Williams, A., Venetis, C. A., Davies, A., Ogden, J., Whelan, L., Hughes, B., Dalton, B., Boy, F., & Chambers, C. (2014) The association between exaggeration in health related science news and academic press releases: retrospective observational study. *BMJ*, 349, g7015
- Tomprowski, P. D., Davis, C. L., Miller, P. H., Naglieri, J. A. (2007) Exercise and Children's Intelligence, Cognition, and Academic Achievement. *Educational Psychology Review* 20(2), 111.
- Weisberg, D. S., Keil, F. C., Goodstein, J., Rawson, E., & Gray, J. R. (2008) The seductive allure of neuroscience explanations. *Journal of Cognitive Neuroscience*, 20(3), 470-477.